Switched Systems

Switched Systems

Edited by
Janusz Kleban

In-Tech
intechweb.org

Published by In-Teh

In-Teh
Olajnica 19/2, 32000 Vukovar, Croatia

www.intechweb.org
Additional copies can be obtained from:
publication@intechweb.org

First published December 2009
Printed in India

Technical Editor: Zeljko Debeljuh

Switched Systems,
Edited by Janusz Kleban

p. cm.
ISBN 978-953-307-018-6

Preface

The theory of switched systems concerns hybrid dynamical systems composed of a family of continuous–time subsystems and a switching law, orchestrating switching between these subsystems. Systems of this class have been identified in many natural and man–made systems, and are recently widely studied by control theorists, computer scientists, practicing engineers and others. Dynamical systems of heterogeneous nature have numerous applications in control of robotic, mechatronic and mechanical systems, gene regulatory networks, automotive industry, aircraft and air traffic control, switching power converters, and also in communication networks, embedded systems, and in many other fields. From the control–theoretic perspective, the most important issues concerning the switched systems are the following: controllability, observability, stabilization and optimal control. A considerable amount of researchers' effort has been put into mathematical, simulational and practical evaluation of these factors.

This book presents selected issues related to switched systems, including practical examples of such systems. It starts with the analysis of stabilization of saturated switching systems and focuses on available results for switching systems being subject to actuator saturations. In Chapter 2 a methodology for robust adaptive control design for a class of switched nonlinear systems is presented. The authors show that it is possible to obtain a separation between robust stability and robust performance, and clear guidelines for performance optimization via ISS bounds. Chapter 3 is devoted to the design of stabilizing feedback controllers for state–dependent switched nonlinear control system, based on a switched control Lyapunov function approach. The robust H_∞ control for linear switched systems with time delay is studied in Chapter 4. The application of the Linear Parameter Varying method for the design of integrated vehicle control systems, in which several active components are used in co-operation, is discussed in Chapter 5. Chapters 6-8 are related to issues concerning communication systems. The analysis of non–exhaustive cyclic service systems with finite capacity, using state space modeling technique, is performed in Chapter 6. A switching network for packet network node and packet dispatching algorithms are presented in Chapter 7, while the modeling and analysis of reliability, availability and serviceability, against hardware faults, for the Sun Datacenter Switch 3456 system is proposed in Chapter 8.

This book is intended for people interested in switched systems, especially researchers and engineers. Graduate and undergraduate students in the area of switched systems can find this book useful to broaden their knowledge concerning control and switching systems. I would like to thank all scientists who have contributed to this book. Gratitude should be shown also to the team at InTech for the initiative and help in publishing this anticipated book.

Editor
Janusz Kleban
Poznan University of Technology
Poland

Contents

1

Stabilization of saturated switching systems

Abdellah Benzaouia
University of Cadi Ayyad, Faculty of Science Semlalia
BP 2390, Marrakech, Morocco
benzaouia@ucam.ac.ma

1. INTRODUCTION

Switched systems are a class of hybrid systems encountered in many practical situations which involve switching between several subsystems depending on various factors. Generally, a switching system consists of a family of continuous-time subsystems and a rule that supervises the switching between them. This class of systems have numerous applications in the control of mechanical systems, the automotive industry, aircraft and air traffic control, switching power converters and many other fields. Two main problems are widely studied in the literature according to the classification given in (Blanchini and Savorgnan, 2006): The first one, which is the one solved in this work, looks for testable conditions that guarantee the asymptotic stability of a switching system under arbitrary switching rules, while the second is to determine a switching sequence that renders the switched system asymptotically stable (see (Liberzon and Morse, 1999) and the reference therein). Following the first approach, (Blanchini *et al.,* 2009) investigate the problem of designing a switching compensator for a plant switching amongst a (finite) family of given configurations (A_i, B_i, C_i).
A main problem which is always inherent to all dynamical systems is the presence of actuator saturations. Even for linear systems, this problem has been an active area of research for many years. Besides approaches using anti-windup techniques (Mulder *et al.,* 2004) and model predictive controls (Camacho and Bordons, 2004), two main approaches have been developed in the literature: The first is the so-called positive invariance approach which is based on the design of controllers which work inside a region of linear behavior where saturations do not occur (see (Benzaouia and Burgat, 1988), (Benzaouia and Hmamed, 1993), (Blanchini, 1999) and the references therein). This approach has already being applied to a class of hybrid systems involving jumping parameters (Benzaouia and Boukas, 2002). It has also been used to design controllers for switching systems with constrained control under complete modelling taking into account reset functions at each switch and different system's dimension . The second approach, however allows saturations to take effect while guaranteeing asymptotic stability (see (Nguyen and Jabbari, 1999, 2000), (Tarbouriech *et al.,* 2006), (Hu *et al.,* 2002)-(Hu and Lin, 2002), (Benzaouia *et al.,* 2006) and the references therein). The main challenge in these two approaches is to obtain large domains of initial states which ensures asymptotic stability for the system despite the presence of saturations (Gilbert and Tan 1991), (Benzaouia and Baddou, 1999), (Benzaouia *et al.,* 2002), (Hu *et al.,* 2002).
The objective of this chapter is to present the available results in the literature for switching systems subject to actuator saturations. These results follow generally two ways: the first concerns the synthesis of non saturating controllers (controllers working inside a large region

of linear behavior where the saturations do not occur), while the second extends the results obtained for unsaturated switching systems by (Mignone *et al.*, 2000), (Ferrari-Trecate *et al.*, 2001) and (Daafouz *et al.*, 2001, 2002) leading to saturating controllers (controllers tolerating saturations to take effect). The second method was firstly used in (Benzaouia *et al.*, 2004) with the use of a multiple Lyapunov function. However, only the intersection of all the corresponding level sets of the local functions was considered as a region of asymptotic stability of the switching system. This drawback is improved in (Benzaouia *et al.*, 2006) and (Benzaouia *et al.*, 2009a) by considering, for the first time, a large set of asymptotic stability composed by the union of all the level sets.
In this context, two main different sufficient conditions of asymptotic stability were obtained for switching systems subject to actuator saturations. Furthermore, these conditions were presented in the form of LMIs for the state feedback control case. A particular attention was given to the output feedback case which has an additive complexity due to the output equation. It was also shown that the LMIs obtained for computing controllers working inside a large region of linear behavior are less conservative.
The obtained results are then extended to uncertain switching system subject to actuator saturations as developed in (Benzaouia *et al.*, 2009b) and (Benzaouia *et al.*, 2009c) respectively. The uncertainty types considered in these two works are the polytopic one and the structured one. This second type of uncertainty was also studied, without saturation, in (Hetel *et al.*, 2006). Thus, in this work (Benzaouia *et al.*, 2009a), two directions are explored: the first concerns the synthesis of non saturating controllers, while the second direction deals with controllers tolerating saturations to take effect under polytopic uncertainties. For structured uncertainties studies in (Benzaouia *et al.*, 2009b), the synthesis of the controller follows two different approaches, the first one deals firstly with the nominal system and then uses a test to check the asymptotic stability in presence of uncertainties while the second considers the global representation of the uncertain system.

2. Stabilization of switching systems subject to actuator saturation

2.1 PROBLEM FORMULATION

In this section, we give a more precise problem statement for the class of systems under consideration, namely, discrete-time switching linear systems with input saturation and state or output feedback. An equivalent description of such systems, based on the indicator function is also used in this work. The main results of this section are published in (Benzaouia *et al.*, 2009a).
Thus, we consider systems described by:

$$\begin{aligned} x(t+1) &= A_\alpha x(t) + B_\alpha sat(u(t)) \\ y(t) &= C_\alpha x(t) \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the control, $y \in \mathbb{R}^p$ is the output, $sat(.)$ is the standard saturation function and α a switching rule which takes its values in the finite set $\mathcal{I} := \{1, \ldots, N\}$, $t \in \mathbb{Z}_+$. The saturation function is assumed here to be normalized, *i. e.*, $(|sat(u)| = min\{1, |u|\})$. Each subsystem α is called a mode.

Definition 2.1. *(Lygeros et al., 1999) An hybrid time basis τ is an infinite or finite sequence of sets $I_s = \{t \in \mathbb{N} : t_s \leq t \leq \bar{t}_s\}$, with $\bar{t}_s = t_{s+1}$ for $s \in \mathcal{L} = \{0, \ldots, L\}$, and if $card(\tau) = L + 1 < \infty$ then $\bar{t}_L$ can be finite or infinite.*

Throughout this chapter, it is assumed that:

- The switching system is stabilizable;
- Matrices C_α are of full rank;
- $t_{s+1} \geq t_s + 1, \forall s \in \mathcal{L}$;
- the switching rule is not known *a priori* but $\alpha(t)$ is available at each t.

The third assumption ensures that at each time only one subsystem is active. The fourth assumption corresponds to practical implementations where the switched system is supervised by a discrete-event system or operator allowing for $\alpha(t)$ to be known in real time.
In this work, we are interested by the synthesis of stabilizing controllers for this class of hybrid systems subject to actuator saturation. We use a feedback control law as:

$$u(t) = F_\alpha x(t) = K_\alpha y(t), \tag{2}$$

and write the closed-loop system as

$$\begin{aligned} x(t+1) &= A_\alpha x(t) + B_\alpha sat(F_\alpha x(t)). \\ &= A_\alpha x(t) + B_\alpha sat(K_\alpha y(t)) \end{aligned} \tag{3}$$

Upon introducing the indicator function:

$$\xi(t) = [\xi_1(t), \ldots, \xi_N(t)]^T \tag{4}$$

where $\xi_i(t) = 1$ if the switching system is in mode i and $\xi_i(t) = 0$ if it is in a different mode, one can write the closed-loop system (3) as follows:

$$\begin{aligned} x(t+1) &= \sum_{i=1}^{N} \xi_i(t)[A_i x(t) + B_i sat(F_i x(t))]. \\ &= \sum_{i=1}^{N} \xi_i(t)[A_i x(t) + B_i sat(K_i C_i x(t))] \end{aligned} \tag{5}$$

2.2 PRELIMINARY RESULTS

In this section, we recall two results on which our work is based. Let α be fixed. Then System (3) becomes a linear time-invariant system with input saturation given by:

$$x(t+1) = Ax(t) + Bsat(Fx(t)) \tag{6}$$

Define the following subsets of $\mathbb{R}^n$:

$$\varepsilon(P, \rho) = \{x \in \mathbb{R}^n | x^T P x \leq \rho\}, \tag{7}$$

$$\mathcal{L}(F) = \{x \in \mathbb{R}^n || F_l x| \leq 1, \, l = 1, \ldots, m\}, \tag{8}$$

with P a positive definite matrix, $\rho > 0$ and F_l the lth row of the matrix $F \in \mathbb{R}^{m \times n}$. Thus $\varepsilon(P, \rho)$ is an ellipsoid while $\mathcal{L}(F)$ is a polyhedral consisting of states for which the saturation does not occur.

Lemma 2.1. *(Hu et al., 2002) For all $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^m$ such that $|v_l| < 1, l \in [1, m]$*

$$sat(u) \in co\{D_s u + D_s^- v, s \in [1, \eta]\} \; ; \;\; \eta = 2^m \tag{9}$$

where co denotes the convex hull.

Consequently, there exist $\delta_1 \geq 0, \ldots, \delta_\eta \geq 0$ with $\sum_{s=1}^{\eta} \delta_s = 1$ such that,

$$sat(u) = \sum_{s=1}^{\eta} \delta_s [D_s u + D_s^- v] \tag{10}$$

Here, D_s is an m by m diagonal matrix with elements either 1 or 0 and $D_s^- = \mathbb{I}_m - D_s$. There are 2^m possible such matrices. One can also consult the work of (Benzaouia *et al.*, 2006) for more details and other extensions to linear systems with both constraints on the control and the increment or rate of the control.
Consider the following autonomous switching system:

$$x(t+1) \quad = \quad \sum_{i=1}^{N} \xi_i(t) A_i x(t) \tag{11}$$

The use of the Lyapunov functions revealed two ways:

- The existence of a common Lyapunov function to the various subsystems guarantees the asymptotic stability of the switching system. In general, the search for such function is not always obvious (Shorten and Narenda, 1997, 1998).
- The multiple Lyapunov functions were introduced in (Branicky, 1998). They are considered as a strong tool in the analysis of the stability of the hybrid systems and in particular the switching systems.

Proposition 2.1. *(Branicky, 1998) If there exists a multiple Lyapunov function $V(t,x), t \in I_s$, with $V : \mathbb{N} \times \mathbb{R}^n \longrightarrow \mathbb{R}_+$, such that the following hold:*

- $\Delta V(t, x(t)) < 0$, *for any time* $t \in I_s$
- $V(t_k, x(t_k)) < V(t_{k-1}, x(t_{k-1})), \forall k \in \mathcal{L}$

then, the switching autonomous system (11) is asymptotically stable.

We now recall a useful stability result for switching systems with no input saturations presented by many authors (see (Mignone *et al.*, 2000), (Ferrari-Trecate *et al.*, 2001) and (Daafouz *et al.*, 2001, 2002)) firstly used for linear time varying systems (Daafouz and Bernussou, 2001).

Theorem 2.1. *The closed-loop switching system (11) is asymptotically stable at the origin if there exist N symmetric and positive definite matrices $P_1, \ldots, P_N$ satisfying,*

$$A_i^T P_j A_i - P_i < 0, \forall (i,j) \in \mathcal{I} \times \mathcal{I} \tag{12}$$

A corresponding Lyapunov function for the system is then given by:

$$V(t,x) = x^T(t)(\sum_{i=1}^{N} \xi_i(t) P_i) x(t) \tag{13}$$

It is worth to note that function $V(t,x)$, which is a multiple Lyapunov function candidate involving matrices P_i, can be seen as a standard Lyapunov function candidate. It was the way followed by (Mignone *et al.*, 2000) and (Daafouz *et al.*, 2001, 2002). Further, condition (12) is equivalent, by using Schur complement to,

$$\left[\begin{array}{cc} P_i & A_i^T P_j \\ * & P_j \end{array} \right] > 0 \ , \ \forall (i,j) \in \mathcal{I} \times \mathcal{I} \tag{14}$$

where $*$ denotes the transpose of the off diagonal element of the LMI. Subsequently, we will need the following equivalent LMI representation of (14):

$$\left[\begin{array}{cc} X_i & X_i A_i^T \\ * & X_j \end{array} \right] > 0 \ , \ \forall (i,j) \in \mathcal{I} \times \mathcal{I} \tag{15}$$

where $X_i = P_i^{-1}$. By noting that inequality (12) is equivalent to:

$$G_i A_i^T P_j A_i G_i^T - G_i P_i G_i^T < 0, \forall (i,j) \in \mathcal{I} \times \mathcal{I} \tag{16}$$

for any nonsingular matrix G_i. By using the fact that $(G_i - X_i)^T X_i^{-1} (G_i - X_i) \geq 0$, implies $G_i X_i^{-1} G_i^T \geq G_i + G_i^T - X_i$. The LMI (15) is also in turn equivalent to the following LMI, generally used to relax the previous one (Daafouz *et al.*, 2001, 2002):

$$\left[\begin{array}{cc} G_i + G_i^T - X_i & G_i^T A_i^T \\ * & X_j \end{array} \right] > 0, \forall (i,j) \in \mathcal{I} \times \mathcal{I} \tag{17}$$

where matrices G_i called slack variables are nonsingular matrices and X_i the positive definite matrices.
It is worth nothing that condition (12) has to be satisfied $\forall (i,j) \in \mathcal{I} \times \mathcal{I}$, in particular for $i = j$. This means that each mode is necessarily asymptotically stable and $V_i(t,x) = x^T(t) P_i x(t)$ is the associated Lyapunov function. Recall that a level set of the Lyapunov function $V(t,x)$ given by (13) and associated to the switching system (11) is given by the set $\varepsilon(P,\rho)$ defined by (7) with $P = \sum_{i=1}^{N} \xi_i(t) P_i$. This region of asymptotic stability is very difficult to construct. Nevertheless, an estimate can be obtained as large as possible as presented in the next section based on the union of the sub-level ellipsoid sets $\varepsilon(P_i, 1)$. Other type of level sets obtained with different Lyapunov functions for switched systems can be found in (Hu *et al.*, 2006). A useful lemma is also recalled.

Lemma 2.2. *Let R, S and Γ be matrices with appropriate dimension. Suppose $\Gamma^T \Gamma \leq \mathbb{I}$, then for any scalar $\lambda > 0$, we have:*

$$R\Gamma S + S^T \Gamma^T R^T \leq \lambda R R^T + \lambda^{-1} S S^T \tag{18}$$

2.3 Analysis and synthesis of stabilizability

In this section, the region of local asymptotic stability associated to the saturated switching system is firstly studied. The design of a stabilizing controller for the class of switching system with actuator saturation is then presented by following two ways, the first concerns controllers working inside the region of linear behavior where the saturations do not occur while the second is based on Lemma 2.1 tolerating saturations to occur.

2.3.1 Region of asymptotic stability

The link between the result of Theorem 2.1 and the level set (7) is not mentioned in (Mignone *et al.*, 2000) nor in (Daafouz *et al.*, 2001, 2002). The absence of saturation on the control of the switching systems in these works does not necessitate to take care with the level sets since the asymptotic stability is global. We present here after an interesting result on this subject inspired from the general class studied in (Benzaouia *et al.*, 2007).

Theorem 2.2. *If there exist N symmetric positive definite matrices $P_1, \ldots, P_N$ such that the inequalities (12) are satisfied then the set given by,*

$$\Omega = \bigcup_{i=1}^{N} \varepsilon(P_i, 1) \tag{19}$$

is a level set of the multiple Lyapunov function $V(t,x)$ given by (13) with respect to the trajectories of the autonomous switching system (11).

Proof: Note that condition (12) for $i = j$ ensures that $\Delta V(t,x) < 0$, for $t \in I_s, s \in \mathcal{L}$ (Daafouz *et al.*, 2001, 2002). Now let $x(\bar{t}_{k-1}) \in \varepsilon(P_i, 1)$. What happens at the switching time $\bar{t}_{k-1}$?
For this, compute $x(t_k)$. With an identity reset function we have,

$$x(t_k) = x(\bar{t}_{k-1}) \tag{20}$$

The switch to the next region $\varepsilon(P_j, 1)$ will be seen an unit time after,

$$\begin{aligned} x(t_k + 1) &= x(\bar{t}_{k-1} + 1) \\ &= A_i x(\bar{t}_{k-1}) \end{aligned} \tag{21}$$

We have,

$$x^T(t_k+1)P_j x(t_k+1) = x^T(\bar{t}_{k-1}) \left(A_i^T P_j A_i \right) x(\bar{t}_{k-1}) \tag{22}$$

According to condition (12), one can obtain,

$$x^T(t_k+1)P_j x(t_k+1) < x^T(\bar{t}_{k-1}) P_i x(\bar{t}_{k-1}) \tag{23}$$

Since $x(\bar{t}_{k-1}) \in \varepsilon(P_i, 1)$, then, $x(t_k + 1) \in \varepsilon(P_j, 1), \forall i, j \in \mathcal{I} \times \mathcal{I}$. This proves that $V(t_k, x(t_k)) < V(t_{k-1}, x(t_{k-1})), \forall k \in \mathcal{L}$. Consequently, function $V(t,x)$ is a multiple Lyapunov function of the switching system according to Proposition 2.1 and the set Ω is a level set associated to this function. It is worth to note that the same reasoning holds for $\varepsilon(P_j, \rho)$ for any positive scalar ρ. □

Note that an important remark is to be done at this level: since the studied system is a switching one composed of N subsystems (modes), the switching system can be initialized inside any level set $\varepsilon(P_i, 1)$, however, the initial mode to be selected is the corresponding mode i.

2.3.2 State feedback control

We assume that the state is available, so a state feedback control can be performed. The first result we present is a simple extension of known result given by (Gutman and Hagandar, 1985) for linear systems and concerns the synthesis of non saturating controllers ensuring that a large region of linear behavior is a region of asymptotic stability.

Theorem 2.3. *If there exists N symmetric matrices $X_1, \ldots, X_N$ and N matrices $Y_1, \ldots, Y_N$ solutions of the following LMIs:*

$$\begin{bmatrix} X_i & (A_i X_i + B_i Y_i)^T \\ * & X_j \end{bmatrix} > 0, \tag{24}$$

$$\begin{bmatrix} 1 & Y_{li} \\ * & X_i \end{bmatrix} > 0, \tag{25}$$

$$\forall (i,j) \in \mathcal{I} \times \mathcal{I}, \forall l \in [1, m]$$

where Y_{li} is the lth row of matrix Y_i; then the switching system with saturations in closed-loop (3), with,

$$F_i = Y_i X_i^{-1}, P_i = X_i^{-1} \tag{26}$$

is asymptotically stable at the origin $\forall x_0 \in \Omega$ and for any sequences of switching $\alpha(t)$.

Proof: In order to guarantee that the control by state feedback is always admissible, each sublevel set of the Lyapunov function has to be contained inside the polyhedral set where the saturations do not occur $\varepsilon(P_i, 1) \subset \mathcal{L}(F_i), \forall i \in \mathcal{I}$ (Gutman and Hagandar, 1985). Using (Boyd *et al.*, 1994), this inclusion condition can also be transformed to the equivalent LMI (25) by letting $X_i = P_i^{-1}$ and $F_i X_i = Y_i$. Let $x(t) \in \Omega, \forall t \in I_s$. According to condition (25), the system in closed-loop (5) can be equivalently written as,

$$x(t+1) \quad = \quad \sum_{i=1}^{N} \xi_i(t) A_{c_i} x(t). \tag{27}$$

with, $A_{c_i} = A_i + B_i F_i$. The proof follows from the asymptotic stability conditions of the switching system given by Theorem 3.2, which is expressed equivalently by (15), with matrix A_{c_i} instead of A_i. Further, Theorem 2.2 ensures that the set Ω is a set of asymptotic stability of the switching system with saturations in closed-loop (3). □

To achieve a domain of attraction as large as possible, we can solve the following optimization problem:

$$(Pb.1) : \begin{cases} \sup_{(X_i, Y_i)} Trace(X_i) \\ s.t. \ \ (24), (25), \\ i = 1, \ldots, N \end{cases}$$

When this optimization problem is feasible, the obtained ellipsoid volumes are maximum with respect to the data of the system.

The obtained LMIs (24) of Theorem 2.1 are similar to those obtained in (Mignone *et al.*, 2000) and (Ferrari-Trecate *et al.*, 2001) for non saturated switching systems. The presence of saturation on the control in our problem leads to the additional LMIs (25), which will obviously restrict the set of solutions. Nevertheless, the associated large region of asymptotic stability Ω for the saturated switching system enables one to conclude that these LMIs are not conservative.

The second result we present concerns the synthesis of saturating controllers tolerating saturations to take effect inside a large region of asymptotic stability.

Theorem 2.4. *If there exist symmetric positive definite matrices $P_1, \ldots, P_N \in \mathbb{R}^{n \times n}$ and matrices $H_1, \ldots, H_N \in \mathbb{R}^{m \times n}$ such that,*

$$\left[\begin{array}{cc} P_i & [A_i + B_i(D_{is} F_i + D_{is}^{-} H_i)]^T P_j \\ * & P_j \end{array} \right] > 0 \ , \quad \forall (i,j) \in \mathcal{I} \times \mathcal{I}, \forall s \in [1, \eta] \tag{28}$$

and,

$$\varepsilon(P_i, 1) \subset \mathcal{L}(H_i), \forall i \in \mathcal{I} \tag{29}$$

then, the closed-loop switching system (3) is asymptotically stable at the origin $\forall x_0 \in \Omega$ and for all sequences of switching $\alpha(t)$.

Proof: Assume that there exist N matrices $H_1, \ldots, H_N$ and N symmetric matrices $P_1, \ldots, P_N$ such that condition (28) and (29) are satisfied. Using the expression in (10) and rewriting System (3) as in (5) yields that:

$$sat(F_i x(t)) = \sum_{s=1}^{\eta} \delta_{is}(t)[D_{is}F_i + D_{is}^{-}H_i]x(t)\ ; \tag{30}$$

$$\delta_{si}(t) \geq 0,\ \sum_{s=1}^{\eta} \delta_{si}(t) = 1 \tag{31}$$

and, subsequently :

$$\begin{aligned} x(t+1) &= \sum_{s=1}^{\eta}\sum_{i=1}^{N} \xi_i(t)\delta_{is}(t)Ac_{is}x(t); \\ Ac_{is} &:= A_i + B_i(D_{is}F_i + D_{is}^{-}H_i),\quad s \in [1,\eta] \end{aligned} \tag{32}$$

The rate of change of the Lyapunov function candidate (13) along the trajectories of (32) is given by:

$$\Delta V(t,x(t)) = x^T(t+1)(\sum_{j=1}^{N} \xi_j(t+1)P_j)x(t+1) - x^T(t)(\sum_{i=1}^{N} \xi_i(t)P_i)x(t) = x^T(t) \tag{33}$$
$$\left[\left(\sum_{s=1}^{\eta}\sum_{i=1}^{N} \xi_i(t)\delta_{is}(t)Ac_{is}\right)^T \left(\sum_{j=1}^{N} \xi_j(t+1)P_j\right) \left(\sum_{s=1}^{\eta}\sum_{i=1}^{N} \xi_i(t)\delta_{is}(t)Ac_{is}\right) - \sum_{i=1}^{N} \xi_i(t)P_i\right] x(t)$$

Let condition (28) be satisfied. At this level, for each i, multiply the $j = 1, \ldots, N$ inequalities (28) by $\xi_j(t+1)$ and sum. Multiply the resulting $i = 1, \ldots, N$ inequalities by $\xi_i(t)$ and sum. Multiply again the resulting $s = 1, \ldots, \eta$ inequalities by $\delta_{is}(t)$ and sum. As $\sum_{i=1}^{N} \xi_i(t) = \sum_{j=1}^{N} \xi_j(t+1) = \sum_{s=1}^{\eta} \delta_{is}(t) = 1$, one gets:

$$\left[\begin{array}{cc} \sum_{i=1}^{N} \xi_i(t)P_i & \Gamma \sum_{j=1}^{N} \xi_j(t+1)P_j \\ * & \sum_{j=1}^{N} \xi_j(t+1)P_j \end{array}\right] > 0, \tag{34}$$

with, $\Gamma = \left[\sum_{s=1}^{\eta}\sum_{i=1}^{N} \xi_i(t)\delta_{is}(t)Ac_{is}\right]^T$. The use of Schur complement allows us to write condition (34) under the equivalent form,

$$\left(\sum_{s=1}^{\eta}\sum_{i=1}^{N} \xi_i(t)\delta_{is}(t)Ac_{is}\right)^T \left(\sum_{j=1}^{N} \xi_j(t+1)P_j\right) \left(\sum_{s=1}^{\eta}\sum_{i=1}^{N} \xi_i(t)\delta_{is}(t)Ac_{is}\right) - \sum_{i=1}^{N} \xi_i(t)P_i < 0, \tag{35}$$
$$\forall (i,j) \in \mathcal{I} \times \mathcal{I}, \forall s \in [1,\eta]$$

Letting λ be the largest eigenvalue among all the above matrices, we obtain that

$$\Delta V(t,x(t)) \leq \lambda x^T(t)x(t) < 0. \tag{36}$$

which ensures the desired result. Noting that condition (28) is also satisfied for $i = j$, this implies that each set $\varepsilon(P_i, 1)$ is a contractively invariant set for the corresponding subsystem.

Further, by taking account of condition (29), one can guarantee that for every $x_o \in \varepsilon(P_i, 1) \subset \mathcal{L}(H_i)$, each subsystem is asymptotically stable at the origin. Besides, following Theorem 2.2, conditions (28)-(29) also allow for a state belonging, before the switch, to a set $\varepsilon(P_i, 1) \subset \mathcal{L}(H_i)$, if a switch occurs at any time t_k, that the switch will handle the state to the desired set $\varepsilon(P_j, 1) \subset \mathcal{L}(H_j)$. That means that the set Ω is a set of asymptotic stability of the switching system. □

This stability result can be used for control synthesis as follows.

Theorem 2.5. *If there exists N symmetric matrices $X_1, \ldots, X_N$ and $2N$ matrices $Y_1, \ldots, Y_N$ and $Z_1, \ldots, Z_N$ solutions of the following LMIs:*

$$\begin{bmatrix} X_i & (A_iX_i + B_iD_{is}Y_i + B_iD_{is}^{-}Z_i)^T \\ * & X_j \end{bmatrix} > 0, \tag{37}$$

$$\begin{bmatrix} 1 & Z_{li} \\ * & X_i \end{bmatrix} > 0, \tag{38}$$

$$\forall (i,j) \in \mathcal{I} \times \mathcal{I}, \forall s \in [1, \eta], \forall l \in [1, m]$$

then the switching system with saturations in closed-loop (3), with,

$$F_i = Y_iX_i^{-1}, H_i = Z_iX_i^{-1}, P_i = X_i^{-1} \tag{39}$$

is asymptotically stable at the origin $\forall x_0 \in \Omega$ and for any sequences of switching $\alpha(t)$.

Proof: The Inequalities (28) can be transformed equivalently by Schur complement to the following:

$$P_j \left[A_i + B_i(D_{is}F_i + D_{is}^{-}H_i)\right] P_i^{-1} \left[A_i + B_i(D_{is}F_i + D_{is}^{-}H_i)\right]^T P_j - P_j < 0 \tag{40}$$

By pre and post-multiplying the latter by P_j^{-1}, it follows:

$$\left[A_i + B_i(D_{is}F_i + D_{is}^{-}H_i)\right] P_i^{-1} \left[A_i + B_i(D_{is}F_i + D_{is}^{-}H_i)\right]^T - P_j^{-1} < 0 \tag{41}$$

Pose $X_i = P_i^{-1}$, inequality (41) can then be rewritten as,

$$\left[A_i + B_i(D_{is}F_i + D_{is}^{-}H_i)\right] X_i \left[A_i + B_i(D_{is}F_i + D_{is}^{-}H_i)\right]^T - X_j < 0,$$

that is

$$\left[A_iX_i + B_i(D_{is}F_iX_i + D_{is}^{-}H_iX_i)\right] X_i^{-1} \left[A_iX_i + B_i(D_{is}F_iX_i + D_{is}^{-}H_iX_i)\right]^T - X_j < 0.$$

The use of the Schur complement a second time leads to:

$$\begin{bmatrix} X_i & (A_iX_i + B_iD_{is}F_iX_i + B_iD_{is}^{-}H_iX_i)^T \\ * & X_j \end{bmatrix} > 0 \tag{42}$$

By letting $F_iX_i = Y_i$ and $H_iX_i = Z_i$, the LMI (37) follows readily. Using (Boyd *et al.*, 1994), the inclusion condition (29) can also be transformed to the equivalent LMI (38). □

To obtain larger ellipsoid domains $\varepsilon(P_i, 1)$, we can apply the following optimization problem:

$$(Pb.2): \begin{cases} \sup_{(X_i, Y_i, Z_i)} Trace(X_i) \\ s.t. \ \ (37), (38), \\ i = 1, \ldots, N \end{cases}$$

When this optimization problem is feasible, the obtained ellipsoid volumes are maximum with respect to the data of the system.

Remark 2.1. *The results presented in this section can be extended easily to level sets $\varepsilon(P_i, \rho_i)$ where the scalars $\rho_i, i = 1, \ldots, N$ can be a priori fixed, by using the variables $X_i = P_i/\rho_i$ in the LMIs (24)-(25) and (37)-(38).*

Commennt 2.1. *The resolution of the LMIs (37) for $s \in [1, \eta]$, may be very restrictive. One can relax this resolution by accepting controllers working in a region of linear behavior where the saturations do not occur. The obtained results in this work, where the switching system is with saturated control, are a set of $(2^m \times N^2 + Nm)$ LMIs for computing stabilizing controllers working inside the saturated regions $\mathcal{L}(H_i)$ or only a set of $(N^2 + Nm)$ LMIs for computing stabilizing controllers working inside regions of linear behavior $\mathcal{L}(F_i)$. Note also that the LMIs of Theorem 2.3 can be obtained from LMIs of Theorem 2.5 by letting $D_{is} = \mathbb{I}$ and $D_{is}^- = 0$.*

Example 2.1. *Consider a numerical switching discrete-time system with saturation specified by the two subsystems:*

$$A_1 = \begin{bmatrix} -0.7 & 1 \\ -0.5 & -1.5 \end{bmatrix}; B_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; A_2 = \begin{bmatrix} 0.9 & -1 \\ 1.7 & -1.5 \end{bmatrix}; B_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

We have to solve 5 LMIs with 4 variables to compute controllers working inside a region of linear behavior. The use of the Matlab LMI Toolbox to check our conditions leads to the following results without using the optimization problem $(Pb.1)$.

$$P_1 = \begin{bmatrix} 7.3328 & 18.7125 \\ 18.7125 & 55.9711 \end{bmatrix}; F_1 = \begin{bmatrix} 1.2244 & 0.7535 \end{bmatrix}; Ac_1 = \begin{bmatrix} 0.5244 & 1.7535 \\ -0.5000 & -1.5000 \end{bmatrix};$$

$$P_2 = \begin{bmatrix} 4.8237 & -5.1001 \\ -5.1001 & 5.6715 \end{bmatrix}; F_2 = \begin{bmatrix} 1.7559 & -1.5699 \end{bmatrix}; Ac_2 = \begin{bmatrix} 0.9000 & -1.0000 \\ -0.0559 & 0.0699 \end{bmatrix};$$

Figure 1 presents the two ellipsoid sets of invariance and contractivity $\varepsilon(P_i, 1)$ together with the sets of linear behavior $\mathcal{L}(F_i)$ computed with LMIs (24)-(25).
For the same example, the use of the Matlab LMI Toolbox to check conditions (37)-(38) leads to the following results by using the optimization problem $(Pb.2)$ with 10 LMIs and 6 variables.

$$P_1 = \begin{bmatrix} 7.4032 & 18.1841 \\ 18.1841 & 56.6345 \end{bmatrix}; H_1 = \begin{bmatrix} 1.2534 & 0.8569 \end{bmatrix}; F_1 = \begin{bmatrix} 1.2641 & 0.8526 \end{bmatrix};$$

$$Ac_1 = \begin{bmatrix} 0.5641 & 1.8526 \\ -0.5000 & -1.5000 \end{bmatrix};$$

$$P_2 = \begin{bmatrix} 4.5202 & -4.3908 \\ -4.3908 & 6.5909 \end{bmatrix}; H_2 = \begin{bmatrix} 1.7693 & -1.5845 \end{bmatrix}; F2 = \begin{bmatrix} 1.7697 & -1.5836 \end{bmatrix};$$

$$Ac_2 = \begin{bmatrix} 0.9000 & -1.0000 \\ -0.0697 & 0.0836 \end{bmatrix};$$

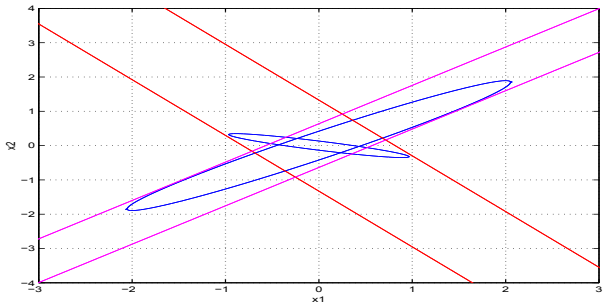


Fig. 1. The ellipsoids sets of invariance and contractivity for the switching discrete-time linear system computed with LMIs (24)-(25).

Note that the optimal values of the optimization problem (Pb.2) are given by $Trace(P_1^{-1}) = 0.7227$ *and* $Trace(P_2^{-1}) = 1.0568$. *Figure 2 presents the two ellipsoid sets of invariance and contractivity*

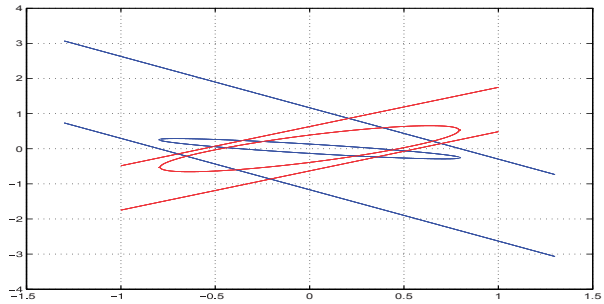


Fig. 2. The ellipsoids sets of invariance and contractivity for the switching discrete-time linear system computed with LMIs (37)-(38) and (*Pb.*2).

$\varepsilon(P_i, 1)$ *together with the set of saturations* $\mathcal{L}(H_i)$ *computed with (Pb.2).*

Comparing the results obtained by LMIs (37)-(38) and LMIs (24)-(25), one can note that the matrices H_i are very closed up to matrices F_i obtained with LMIs (37)-(38), that is, $\mathcal{L}(H_i) \cong \mathcal{L}(F_i)$. This means that in the case of this example, the saturations allowed by this technique are not very important. Recall that inside the sets $\mathcal{L}(F_i)$ no saturations occur. Furthermore, the results obtained with LMIs (24)-(25) are less conservative. This is due to the fact that among the number of LMIs to be solved in (37)-(38), one find the LMIs (24)-(25).

2.3.3 Output feedback control

Static output-feedback control plays a very important role in control applications: The purpose is to design controllers such that the resulting closed-loop system is asymptotically stable without using any reconstruction method of the unavailable states. In this section, we begin by the synthesis of the saturating controllers since the non saturating ones can be derived as a particular case. The synthesis of the stabilizing controller by output feedback control for the class of switching system with actuator saturation is presented by applying

the results of Theorem 2.4.

Theorem 2.6. *If there exist symmetric matrices X_i, matrices V_i, Y_i and Z_i solutions of the following LMIs:*

$$\begin{bmatrix} X_i & (A_iX_i + B_iD_{is}Y_iC_i + B_iD_{is}^{-}Z_i)^T \\ * & X_j \end{bmatrix} > 0, \quad (43)$$

$$\begin{bmatrix} 1 & Z_{il} \\ * & X_i \end{bmatrix} > 0, \quad (44)$$

$$V_iC_i = C_iX_i \quad (45)$$

$$\forall (i,j) \in \mathcal{I}^2, \ \forall s \in [1,\eta], \ \forall l \in [1,m]$$

then the closed-loop saturated switching system (3) with

$$K_i = Y_iV_i^{-1}, H_i = Z_iX_i^{-1}, P_i = X_i^{-1} \quad (46)$$

is asymptotically stable $\forall x_0 \in \Omega$ and for all switching sequences $\alpha(k)$.

Proof: The same development as (40)-(42) while replacing F_i by K_iC_i, can be followed to obtain,

$$\begin{bmatrix} X_i & (A_iX_i + B_iD_{is}K_iC_iX_i + B_iD_{is}^{-}H_iX_i)^T \\ * & X_j \end{bmatrix} > 0.$$

According to Equation (45), we have $K_iC_iX_i = K_iV_iC_i$. By letting $K_iV_i = Y_i$ and $H_iX_i = Z_i$, the LMI (43) follows together with relations (46). Similarly, the inclusion condition (29) can also be transformed to the equivalent LMI (44). Finally, the inequality (43) ensures that the obtained solutions X_i are positive definite while equalities (45) guarantee that matrices V_i are nonsingular. □

It is worth noting that the state feedback follows readily from Theorem 2.6 by letting $C_i = \mathbb{I}_m$. In this case, $V_i = X_i$.

The LMI (43), relating matrices C_i to matrix X_i by means of Equation (45), can be relaxed by using the LMI (17), where the new variables G_i are related to matrices C_i instead of the matrix X_i.

Corollary 2.1. *If there exist symmetric matrices X_i, matrices G_i, V_i, Y_i and Z_i solutions of the following LMIs:*

$$\begin{bmatrix} G_i + G_i^T - X_i & \Psi_{is}^T \\ * & X_j \end{bmatrix} > 0, \quad (47)$$

$$\begin{bmatrix} 1, & Z_{il} \\ * & G_i + G_i^T - X_i \end{bmatrix} > 0, \quad (48)$$

$$V_iC_i = C_iG_i \quad (49)$$

$$\forall (i,j) \in \mathcal{I}^2, \ \forall s \in [1,\eta], \ \forall l \in [1,m]$$

where $\Psi_{is} = (A_iG_i + B_iD_{is}Y_iC_i + B_iD_{is}^{-}Z_i)$, then the closed-loop saturated switching system (3) with

$$K_i = Y_iV_i^{-1}, H_i = Z_iG_i^{-1}, P_i = X_i^{-1} \quad (50)$$

is asymptotically stable $\forall x_0 \in \Omega$ and all switching sequences $\alpha(k)$.

Proof: The proof uses the equivalent LMI (17) and is similar to that of Theorem 2.6. However, the inclusion condition $\varepsilon(P_i,1) \subset \mathcal{L}(H_i) \quad \forall i \in \mathcal{I}$ holds if $1 - H_{il} X_i H_{il}^T > 0, \forall l \in [1,m]$, which is equivalent to,
$1 - (H_i G_i)_l (G_i^T X_i^{-1} G_i)^{-1} (H_i G_i)_l^T > 0$. That is, by virtue of (50) $1 - (Z_{il})(G_i^T X_i^{-1} G_i)^{-1} (Z_{il})^T > 0$. Since, $(G_i - X_i)^T X_i^{-1} (G_i - X_i) \geq 0$, then $G_i X_i^{-1} G_i^T \geq G_i + G_i^T - X_i$. It follows that, $1 - Z_{il}(G_i + G_i^T - X_i) Z_{il}^T > 0$ sufficient to have $\varepsilon(P_i,1) \subset \mathcal{L}(H_i)$. By Schur complement, the LMI (48) is obtained. □
To achieve a domain of attraction as large as possible, we can solve similar optimization problem as for state feedback control.
Another way to deal with the problem by relaxing the more constraining relations (45) of Theorem 2.6, especially in MIMO case, is presented by the following result which uses the same idea as in (El Ghaoui *et al.*, 1997) and (Chadli *et al.*, 2002).

Theorem 2.7. *If there exist symmetric matrices X_i and S_i, matrices Y_i and Z_i solutions of the following LMIs:*

$$\begin{bmatrix} X_i & (A_i + B_i D_{is} Y_i C_i + B_i D_{is}^- Z_i)^T \\ * & S_j \end{bmatrix} > 0, \tag{51}$$

$$\begin{bmatrix} 1 & Z_{il} \\ * & X_i \end{bmatrix} > 0, \tag{52}$$

$$\begin{bmatrix} X_i, & \mathbb{I} \\ * & S_i \end{bmatrix} \geq 0 \tag{53}$$

$$\forall (i,j) \in \mathcal{I}^2, \ \forall s \in [1,\eta], \ \forall l \in [1,m]$$

such that $Trace(X_i S_i) = n$, then the closed-loop saturated switching system (3) with

$$K_i = Y_i, H_i = Z_i, P_i = X_i \tag{54}$$

is asymptotically stable $\forall x_0 \in \Omega$ and all switching sequences $\alpha(k)$.

Proof: The inequality (28) can be equivalently transformed via the Schur Complement to the following:

$$\left[A_i + B_i(D_{is} K_i C_i + D_{is}^- H_i)\right]^T P_j \left[A_i + B_i(D_{is} K_i C_i + D_{is}^- H_i)\right] - P_i < 0 \tag{55}$$

The use of the Schur complement a second time leads to

$$\begin{bmatrix} P_i & (A_i + B_i D_{is} K_i C_i + B_i D_{is}^- H_i)^T \\ * & P_j^{-1} \end{bmatrix} > 0 \tag{56}$$

By letting $X_i = P_i$, $S_j = P_j^{-1}$, $Y_i = K_i$ and $Z_i = H_i$, the LMI (51) follows together with relations (54). The inclusion condition (29) can also be transformed to the equivalent LMI (52) (Boyd *et al.*, 1994): $H_{il} P_i^{-1} H_{il}^T \leq 1$ can be rewritten by Schur complement, while using $Z_i = H_i$ as (52). Note that the LMI (51) ensures that the obtained solutions X_i and S_i are positive definite. Finally, the LMI (53) is equivalent to $S_i X_i \geq \mathbb{I}$. □

This result presents the advantage of computing directly the matrices K_i and H_i. Nevertheless, an optimization problem must be solved to achieve $X_i S_i \simeq \mathbb{I}$ by minimizing iteratively the trace of matrix $X_i S_i$ using an algorithm presented in (El Ghaoui *et al.*, 1997). This heuristic is based on a linear approximation of $Tr(X_i S_i)$ by $Tr(X_0 S_i + S_0 X_i)$ where X_0 and S_0 are particular solutions of the LMI constraints (51), (52) and (53).

$$(Pb.3): \left\{ \begin{array}{c} \min_{(S_j, X_i, Y_i, Z_i)} Trace(X_i S_i) \\ s.t. \ \ (51), (52), (53) \\ i, j = 1, \ldots, N \end{array} \right.$$

An output feedback non saturating controller can be obtained, as noted in Comment 2.1, as a particular case of the saturating controller results by letting $D_{is} = \mathbb{I}$ and $D_{is}^{-} = 0$. It is also obvious that the inclusion condition becomes $\varepsilon(P_i, 1) \subset \mathcal{L}(K_i C_i)$. The following results are then directly obtained.

Corollary 2.2. *If there exist symmetric matrices X_i, matrices V_i and Y_i solutions of the following LMIs:*

$$\left[\begin{array}{cc} X_i & (A_i X_i + B_i Y_i C_i)^T \\ * & X_j \end{array} \right] > 0, \quad (57)$$

$$\left[\begin{array}{cc} 1 & (Y_i C_i)_l \\ * & X_i \end{array} \right] > 0, \quad (58)$$

$$V_i C_i = C_i X_i \quad (59)$$

$$\forall (i, j) \in \mathcal{I}^2, \forall l \in [1, m]$$

then the closed-loop saturated switching system (3) with

$$K_i = Y_i V_i^{-1}, P_i = X_i^{-1} \quad (60)$$

is asymptotically stable $\forall x_0 \in \Omega$ and for all switching sequences $\alpha(k)$.

Proof: This proof can be obtained easily by letting $D_{is} = \mathbb{I}$ and $D_{is}^{-} = 0$ in Theorem 2.6. However, the inclusion condition $\varepsilon(P_i, 1) \subset \mathcal{L}(K_i C_i)$ leads to $(K_i C_i)_l P_i^{-1} (K_i C_i)_l^T \leq 1$ which is equivalent to,

$$(K_i C_i P_i^{-1})_l P_i (K_i C_i P_i^{-1})_l^T \leq 1,$$

by using (59), $(K_i V_i C_i)_l X_i^{-1} (K_i V_i C_i)_l^T \leq 1$, can be rewritten by Schur complement, while using $Y_i = K_i V_i$ as (58). □

In a similar way, the result of Theorem 2.7 can be applied for designing non saturating controls by letting $D_{is} = \mathbb{I}$ and $D_{is}^{-} = 0$.

Example 2.2. *In order to illustrate these results, consider a SISO saturated switching discrete-time system with two modes given by the following matrices:*

$$A_1 = \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \right]; B_1 = \left[\begin{array}{c} 10 \\ 5 \end{array} \right]; C_1 = \left[\begin{array}{cc} 1 & 1 \end{array} \right]$$

$$A_2 = \left[\begin{array}{cc} 0 & -1 \\ 0.0001 & 1 \end{array} \right]; B_2 = \left[\begin{array}{c} 0.5 \\ -2 \end{array} \right]; C_2 = \left[\begin{array}{cc} 2 & 3 \end{array} \right]$$

For this example with $n = 2$, $m = p = 1$ and $N = 2$, we have to solve 13 LMIs with 9 variables by using the LMIs of Corollary 2.1. Let the scalar $\rho = 1$. The use of the Matlab LMI Toolbox yields the following results:

$$P_1 = \begin{bmatrix} 0.0016 & 0.0006 \\ 0.0006 & 0.0025 \end{bmatrix}; H_1 = \begin{bmatrix} -0.0196 & -0.0479 \end{bmatrix};$$
$$K_1 = -0.0743; Ac_1 = \begin{bmatrix} 0.2574 & 0.2574 \\ -0.3713 & 0.6287 \end{bmatrix};$$

$$P_2 = \begin{bmatrix} 0.0010 & 0.0004 \\ 0.0004 & 0.0027 \end{bmatrix}; H_2 = \begin{bmatrix} 0.000007 & 0.0483 \end{bmatrix};$$
$$K_2 = 0.0347; Ac_2 = \begin{bmatrix} 0.0615 & -0.9077 \\ -0.2460 & 0.6309 \end{bmatrix}$$

The main conditions (12) are also satisfied:

$$Ac_1^T P_2 Ac_1 - P_1 = 10^{-3} \begin{bmatrix} -0.0012 & -0.0011 \\ -0.0011 & -0.0013 \end{bmatrix};$$
$$Ac_2^T P_1 Ac_2 - P_2 = 10^{-3} \begin{bmatrix} -0.0008 & -0.0007 \\ -0.0007 & -0.0011 \end{bmatrix}$$

Figure 3 shows the level set as the union of two ellipsoid sets of invariance and contractivity for

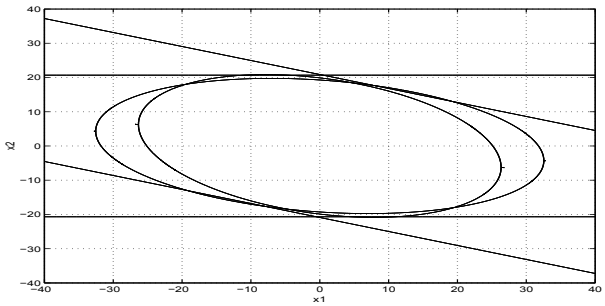


Fig. 3. The level set as the union of two ellipsoid sets of invariance and contractivity for the switching discrete-time linear system obtained by Corollary 2.1.

the switching discrete-time linear system obtained by Corollary 2.1. Each ellipsoid set is contained inside the set of admissible saturations $\mathcal{L}(H_i)$. Inside this set, the asymptotic stability of the saturated switching system is guaranteed for any arbitrary switch from any linear subsystem to another. This SISO example is studied only in order to illustrate the results of this work by plotting the level set in the plane. The use of Theorem 2.7 leads to the following results for $\rho = 1$:

$$P_1 = \begin{bmatrix} 1.0961 & 0.0848 \\ 0.0848 & 1.0290 \end{bmatrix}; P_2 = \begin{bmatrix} 1.5366 & 0.1255 \\ 0.1255 & 1.4821 \end{bmatrix};$$

$$S_1 = \begin{bmatrix} 0.6553 & -0.0555 \\ -0.0555 & 0.6794 \end{bmatrix}; S_2 = \begin{bmatrix} 0.9182 & -0.0756 \\ -0.0756 & 0.9781 \end{bmatrix};$$

$$H_1 = \begin{bmatrix} -0.0786 & -0.1208 \end{bmatrix}; K_1 == -0.1007; H_2 = \begin{bmatrix} 0.00004 & 0.5473 \end{bmatrix}; K2 = 0.1812$$

The level set obtained with this approach is plotted in Figure 4 together with few trajectories.

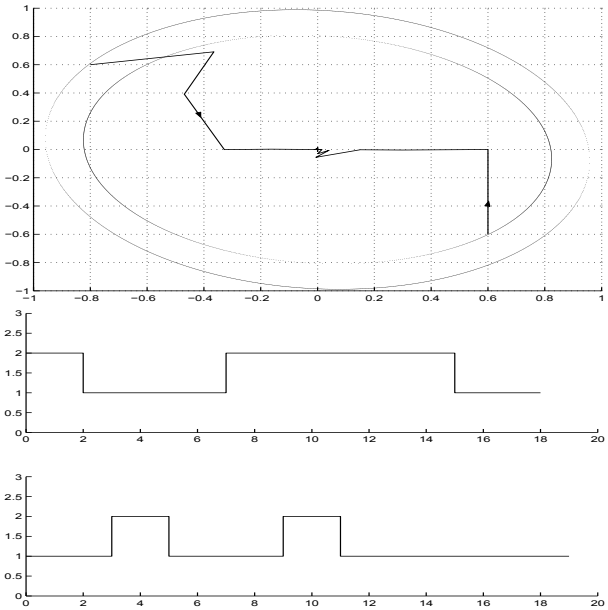


Fig. 4. The level set as the union of two ellipsoid sets of invariance and contractivity for the switching discrete-time linear system obtained by Theorem 2.7 and the corresponding switching sequences.

One can notice that the obtained level sets with Corollary2.1 are the most larger. One can conclude, via this example, that the result of Corollary 2.1 is the less conservative among all the other results due to the introduced slack variables G_i. Combining the comparisons made for state feedback control and output feedback control, one can expect that the introduction of slack variables in the Corollary 2.2 can lead to the more less conservative results.
In this section, two main different sufficient conditions of asymptotic stability are obtained for switching discrete-time linear systems subject to actuator saturations for each case: state feedback and output feedback control. The first allows the synthesis of stabilizing controllers inside a large region of linear behavior while the second applies the idea of Lemma 2.1 which rewrites the saturation function under a combination of 2^m elements to obtain stabilizing controllers tolerating saturations to take effect.

3. Stabilization of saturated switching systems with polytopic uncertainties

The objective of this section is to extend the results of (Benzaouia *et al.*, 2006) to uncertain switching system subject to actuator saturations. The uncertainty type considered in this work is the polytopic one. This type of uncertainty was also studied, without saturation, in (Hetel *et al.*, 2006). Thus, in this work, two directions are explored: the first concerns the synthesis of non saturating controllers, while the second direction deals with controllers

tolerating saturations to take effect under polytopic uncertainties. The main results of this section are published in (Benzaouia *et al.*, 2009b).

3.1 Problem presentation

Let us consider the uncertain saturated switching discrete-time linear system described by:

$$\begin{cases} x_{t+1} = A_\alpha(q_\alpha(t))x_t + B_\alpha(q_\alpha(t))sat(u_t) \\ y_t = C_\alpha(q_\alpha(t))x_t \end{cases} \tag{61}$$

where $x_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^m$, $y_t \in \mathbb{R}^p$ are the state, the input and the output respectively, $sat(.)$ is the standard saturation (assumed here to be normalized, *i.e.*, $|sat(u_t)| = min(1, |u_t|)$), function $\alpha(t) : \mathbb{N} \longmapsto \mathcal{I}$ is a switching rule taking its values $\alpha(t) = i$ in the finite set $\mathcal{I} = \{1, ..., N\}$ and $q_\alpha(t) \in \Gamma_\alpha \subset \mathbb{R}^{d_\alpha}$ are the bounded uncertainties that affect the system parameters in such a way that

$$A_\alpha(q_\alpha(t)) = A_\alpha + \sum_{h=1}^{d_\alpha} A_{\alpha h} q_{\alpha h}(t) \tag{62}$$

$$B_\alpha(q_\alpha(t)) = B_\alpha + \sum_{h=1}^{d_\alpha} B_{\alpha h} q_{\alpha h}(t) \tag{63}$$

$$C_\alpha(q_\alpha(t)) = C_\alpha + \sum_{h=1}^{d_\alpha} C_{\alpha h} q_{\alpha h}(t) \tag{64}$$

where matrices A_α, B_α, C_α represent the nominal matrices and $q_{\alpha h}(t)$ the hth component of vector $q_\alpha(t)$:

$$q_\alpha(t) = [q_{\alpha 1}(t) \;\; q_{\alpha 2}(t) \;\; ... \;\; q_{\alpha h}(t) \;\; ... \;\; q_{\alpha d_\alpha}(t)]^T.$$

The following additional assumption is required:

- Γ_α are compact convex sets.

Let the control be obtained by an output feedback control law:

$$\begin{aligned} u_t &= K_\alpha y_t = F_\alpha x_t, \\ F_\alpha &= K_\alpha C_\alpha. \end{aligned} \tag{65}$$

Thus, the closed-loop system is given by:

$$x_{t+1} = A_\alpha(q_\alpha(t))x_t + B_\alpha(q_\alpha(t))sat(K_\alpha C_\alpha(q_\alpha(t))x_t). \tag{66}$$

Defining the indicator function:

$$\xi(t) := [\xi_1(t), ..., \xi_N(t)]^T, \tag{67}$$

where $\xi_i(t) = 1$ if the switching system is in mode i and 0 otherwise, yields the following representation for the closed-loop system:

$$x_{t+1} = \sum_{i=1}^{N} \xi_i(t)[A_i(q_i(t))x_t + B_i(q_i(t))sat(K_i C_i(q_i(t))x_t]. \tag{68}$$

Let each convex set Γ_i has μ_i vertices $\nu_{i\kappa}, \kappa = 1, \ldots, \mu_i$ so that for every $q_i \in \Gamma_i$, one can write $q_i = \sum_{\kappa=1}^{\mu_i} \beta_{i\kappa}\nu_{i\kappa}$ with $\sum_{\kappa=1}^{\mu_i} \beta_{i\kappa} = 1, 0 \leq \beta_{i\kappa} \leq 1$. The consequence of this, is that each matrix $A_i(q_i(t))$, $B_i(q_i(t))$ and $C_i(q_i(t))$ can be expressed as a convex combination of the corresponding vertices of the compact set Γ_i as follows:

$$\begin{aligned} M(q_i): &= M_i + \sum_{\kappa=1}^{\mu_i} \beta_{i\kappa} M(\nu_{i\kappa}) = \sum_{\kappa=1}^{\mu_i} \beta_{i\kappa} M_{i\kappa}, \\ M(\nu_{i\kappa}) &= \sum_{h=1}^{d_i} M_{ih}\nu_{i\kappa h}, M_{i\kappa} = M_i + M(\nu_{i\kappa}), \sum_{\kappa=1}^{\mu_i} \beta_{i\kappa} = 1, 0 \leq \beta_{i\kappa} \leq 1. \end{aligned}$$

where M_i represents the nominal matrix. Matrix M can be taken differently as A, B or C. Note that the system without uncertainties can be obtained as a particular case of this representation by letting the vertices $\nu_{i\kappa} = 0, \forall i, \forall \kappa$. Besides, equations (69) are directly related to the dimension d_i of the convex compact set Γ_i. The saturated uncertain switching system given by (68) can be rewritten as:

$$x_{t+1} = \sum_{i=1}^{N} \sum_{\kappa=1}^{\mu_i} \xi_i(t)\beta_{i\kappa}(t)[A_{i\kappa}x_t + B_{i\kappa} sat(K_i C_{i\kappa} x_t)] \tag{69}$$

The nominal matrices will be represented by A_i, B_i and C_i. The nominal system in closed-loop is then given by:

$$x_{t+1} = \sum_{i=1}^{N} \xi_i(t)[A_i x_t + B_i sat(K_i C_i x_t)] \tag{70}$$

3.2 Analysis and synthesis of stabilizability

This section presents sufficient conditions of asymptotic stability of the saturated uncertain switching system given by (69). The synthesis of the controller follows two different approaches, the first one deals firstly with the nominal system and then uses a test to check the asymptotic stability in presence of uncertainties while the second considers the global representation of the uncertain system (69).

Theorem 3.1. *If there exist symmetric positive definite matrices $P_1, \ldots, P_N \in \mathbb{R}^{n\times n}$ and matrices $H_1, \ldots, H_N \in \mathbb{R}^{m\times n}$ such that*

$$\begin{bmatrix} P_i & [A_{i\kappa} + B_{i\kappa}(D_{is}K_i C_{i\kappa} + D_{is}^{-} H_i)]^T P_j \\ * & P_j \end{bmatrix} > 0, \tag{71}$$

$$\forall \kappa = 1, \ldots, \mu_i, \forall (i,j) \in \mathcal{I}^2, \forall s \in [1,\ \eta], \tag{72}$$

and

$$\varepsilon(P_i,\ 1) \subset \mathcal{L}(H_i), \tag{73}$$

then the closed-loop uncertain saturated switching system (69) is asymptotically stable $\forall\ x_0 \in \Omega := \bigcup_{i=1}^{N} \varepsilon(P_i,\ 1)$ and for all switching sequences $\alpha(t)$.

Proof: By using Lemma (2.1), for all $H_i \in \mathbb{R}^{m\times n}$ with $|H_{ij}x_t| < 1$, $j \in [1, m]$, where H_{ij} denotes the jth row of matrix H_i, there exist $\delta_{i\kappa 1} \geq 0$,..., $\delta_{i\kappa\eta} \geq 0$ such that $sat(K_i C_{i\kappa} x_t) =$

$\sum_{s=1}^{\eta} \delta_{is\kappa}(t)[D_{is}K_iC_{i\kappa} + D_{is}^{-}H_i]x_t$, $\delta_{i\kappa s}(t) \geq 0$, $\sum_{s=1}^{\eta} \delta_{i\kappa s}(t) = 1$. Then the closed-loop system (69) can be rewritten as

$$x_{t+1} = \sum_{s=1}^{\eta} \sum_{i=1}^{N} \sum_{\kappa=1}^{\mu_i} \xi_i(t)\beta_{i\kappa}(t)\delta_{i\kappa s}(t)Ac_{i\kappa s}x_t \tag{74}$$

$$Ac_{i\kappa s} := A_{i\kappa} + B_{i\kappa}(D_{is}K_iC_{i\kappa} + D_{is}^{-}H_i).$$

Consider the Lyapunov function candidate $V(x) = x_t^T(\sum_{i=1}^{N} \xi_i(t)P_i)x_t$. Computing its rate of increase along the trajectories of system (69) yields.

$$\begin{aligned}\Delta V(x_t) &= x_{t+1}^T(\sum_{j=1}^{N} \xi_j(t+1)P_j)x_{t+1} - x_t^T(\sum_{i=1}^{N} \xi_i(t)P_i)x_t \\ &= x_t^T \left\{ \Sigma^T(\sum_{j=1}^{N} \xi_j(t+1)P_j)\Sigma - \sum_{i=1}^{N} \xi_i(t)P_i \right\} x_t.\end{aligned}$$

where,

$$\Sigma = \sum_{s=1}^{\eta} \sum_{i=1}^{N} \sum_{\kappa=1}^{\mu_i} \xi_i(t)\beta_{i\kappa}(t)\delta_{s\kappa i}(t)Ac_{i\kappa s}$$

Let condition (71) be satisfied. For each i and j multiply successively by $\xi_i(t)$, $\xi_j(t+1)$, $\beta_{i\kappa}(t)$ and $\delta_{i\kappa s}(t)$ and sum. As $\sum_{i=1}^{N} \xi_i(t) = \sum_{j=1}^{N} \xi_j(t+1) = \sum_{s=1}^{\eta} \delta_{i\kappa s}(t) = \sum_{\kappa=1}^{\mu_i} \beta_{i\kappa}(t) = 1$, one gets:

$$\begin{bmatrix} \sum_{i=1}^{N} \xi_i(t)P_i & \Pi \\ * & \sum_{j=1}^{N} \xi_j(t+1)P_j \end{bmatrix} > 0, \tag{75}$$

where

$$\Pi = \Sigma^T(\sum_{j=1}^{N} \xi_j(t+1)P_j).$$

Inequality (75) is equivalent, by Schur complement, to

$$\Sigma^T(\sum_{j=1}^{N} \xi_j(t+1)P_j)\Sigma - \sum_{i=1}^{N} \xi_i(t)P_i < 0$$

Letting λ be the largest eigenvalue among all the above matrices, we obtain that

$$\Delta V(x_t) \leq \lambda x_t^T x_t < 0, \tag{76}$$

which ensures the desired result. Besides, following Theorem 2.4, (71)-(73) also allow for a state belonging to a set $\varepsilon(P_i, 1) \subset \mathcal{L}(H_i)$, before the switch, if a switch occurs at time t_k, the switch will drive the state to the desired set $\varepsilon(P_j, 1) \subset \mathcal{L}(H_j)$. That means that the set Ω is a set of asymptotic stability of the uncertain saturated switching system. □

Remark 3.1. *It is worth to note that the result of Theorem 2.4 can be obtained as a particular case of Theorem 3.1.*

This stability result is now used for control synthesis in two ways: the first consists in computing the controllers only with the nominal system and to test their robustness in a second step; while the second consists in computing in a single step the robust controllers. At this end, the result of Theorem 2.6 can be used to compute matrices K_i, H_i and P_i for the nominal switching system (70). At this step, the stabilizing controllers K_i and H_i of the nominal system are assumed to be known. Then, the following test has to be performed.

Corollary 3.1. *If there exist symmetric positive definite matrices X_i such that*

$$\left[\begin{array}{cc} X_i & (A_{i\kappa}X_i + B_{i\kappa}D_{is}K_iC_{i\kappa}X_i + B_{i\kappa}D_{is}^{-}H_iX_i)^T \\ * & X_j \end{array} \right] > 0, \tag{77}$$

$$\left[\begin{array}{cc} 1 & (H_iX_i)_l \\ * & X_i \end{array} \right] > 0, \tag{78}$$

$\forall (i,j) \in \mathcal{I}^2, \quad \forall s \in [1,\ \eta], \forall l \in [1,m],\ \forall \kappa \in [1,\ \mu_i]$,
with $P_i = X_i^{-1}$, then the closed loop uncertain switching system (69) is asymptotically stable $\forall\ x_0 \in \bigcup_{i=1}^{N} \varepsilon(P_i,\ 1)$ and for all switching sequences $\alpha(t)$.

Proof: The proof is similar to that given for Theorem 2.6. □

The second way to deal with robust controller design is to run a global set of LMIs leading, if it is feasible, to the robust controllers directly. However, one can note that this method is computationally more intensive.

Theorem 3.2. *If there exist symmetric positive definite matrices X_i, matrices, Y_i, V_i and Z_i such that*

$$\left[\begin{array}{cc} X_i & (A_{i\kappa}X_i + B_{i\kappa}D_{is}Y_iC_{i\kappa} + B_{i\kappa}D_{is}^{-}Z_i)^T \\ * & X_j \end{array} \right] > 0, \tag{79}$$

$$\left[\begin{array}{cc} 1 & Z_{il} \\ * & X_i \end{array} \right] > 0, \tag{80}$$

$$V_iC_{i\kappa} = C_{i\kappa}X_i, \tag{81}$$

$\forall (i,j) \in \mathcal{I}^2, \quad \forall s \in [1,\ \eta], \forall l \in [1,m], \forall \kappa \in [1,\ \mu_i]$
with

$$H_i = Z_iX_i^{-1}, K_i = Y_iV_i^{-1}, P_i = X_i^{-1}, \tag{82}$$

then, the closed-loop uncertain saturated switching system (69) is asymptotically stable $\forall\ x_0 \in \Omega$, and for all switching sequences $\alpha(t)$.

Proof: The proof is also similar to that given for Theorem 2.6. □

In order to relax the previous LMIs, one can introduce some slack variables as in (Daafouz *et al.*, 2002) and (Benzaouia *et al.*, 2006), as it is now shown:

Theorem 3.3. *If there exist symmetric positive definite matrices X_i, matrices, Y_i, V_i, G_i and Z_i such that*

$$\left[\begin{array}{cc} G_i + G_i^T - X_i & \Psi \\ * & X_j \end{array} \right] > 0, \tag{83}$$

with $\Psi = (A_{i\kappa}G_i + B_{i\kappa}D_{is}Y_iC_{i\kappa} + B_{i\kappa}D_{is}^{-}Z_i)^T$,

$$\begin{bmatrix} 1 & Z_{il} \\ * & G_i + G_i^T - X_i \end{bmatrix} > 0, \tag{84}$$

$$V_iC_{i\kappa} = C_{i\kappa}G_i, \tag{85}$$

$\forall \kappa = 1, \ldots, \mu_i, \forall (i,j) \in \mathcal{I}^2, \forall s \in [1, \eta], \forall l \in [1, m]$, *with*

$$H_i = Z_iG_i^{-1}, K_i = Y_iV_i^{-1}, P_i = X_i^{-1}; \tag{86}$$

then, the closed-loop uncertain saturated switching system (69) is asymptotically stable $\forall\ x_0 \in \Omega$ *and for all switching sequences* $\alpha(t)$.

Proof: The proof is similar to that given for Corollary 2.1 □
These results can be illustrated with the following example.

Example 3.1. *Consider a SISO saturated switching discrete system with two modes given by the following matrices:*

$$\begin{aligned} A_1(q_1(t)) &= \begin{bmatrix} 1 & 1 \\ 0 & 1+q_{11} \end{bmatrix}; B_1(q_1(t)) = \begin{bmatrix} 10 \\ 5 \end{bmatrix}; \\ C_1(q_1(t)) &= \begin{bmatrix} 1+q_{12} & 1 \end{bmatrix}; \end{aligned}$$

$$\begin{aligned} A_2(q_2(t)) &= \begin{bmatrix} 0+q_{21} & -1 \\ 0.0001 & 1 \end{bmatrix}; B_2(q_2(t)) = \begin{bmatrix} 0.5 \\ -2+q_{22} \end{bmatrix}; \\ C_2(q_2(t)) &= \begin{bmatrix} 2 & 3 \end{bmatrix}. \end{aligned}$$

The vertices of the domain of uncertainties that affect the first mode are:

$$\begin{aligned} \nu_{11} &= (-0.1, -0.2),\ \ \nu_{12} = (-0.1,\ 0.2) \\ \nu_{13} &= (0.1, -0.2),\ \ \nu_{14} = (0.1, 0.2). \end{aligned}$$

The vertices of the domain of uncertainties that affect the second mode are:

$$\begin{aligned} \nu_{21} &= (-0.2, 0.5),\ \ \nu_{22} = (-0.2,\ -0.1) \\ \nu_{23} &= (0.3, 0.5),\ \ \nu_{24} = (0.3, -0.1). \end{aligned}$$

Using Theorem 2.6, a stabilizing controller for the nominal system is

$$K_1 = -0.1000, \quad K_2 = 0.1622.$$

To test the robustness, we can use the Corollary 3.1 which leads to the following results:

$$P_1 = \begin{bmatrix} 0.0208 & -0.0133 \\ -0.0133 & 0.0257 \end{bmatrix}; P_2 = \begin{bmatrix} 0.0320 & 0.0023 \\ 0.0023 & 0.0474 \end{bmatrix}$$

On the other hand, the use of Theorem 3.2 leads to the following results:

$$K_1 = -0.0902, \quad K_2 = 0.1858.$$

Figures 5, 6 and 7 concern the first method. In Figure 5, the switching signals $\alpha(t)$ and the evolution of uncertainties used for simulation, are plotted. Figure 6 shows the obtained level set of stability $\bigcup_{i=1}^{N} \varepsilon(P_i,\ 1)$ which is well contained inside the sets of saturations, while Figure 7 presents some system motions evolving inside the level set starting from different initial states.

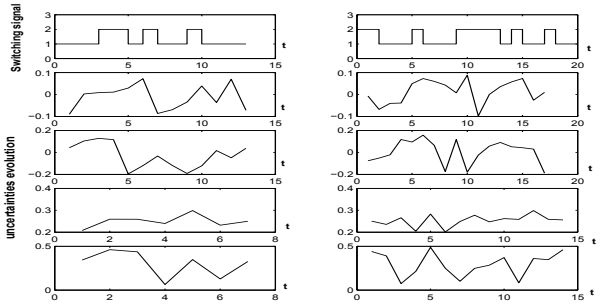


Fig. 5. Switching signals $\alpha(t)$ and uncertainties evolution

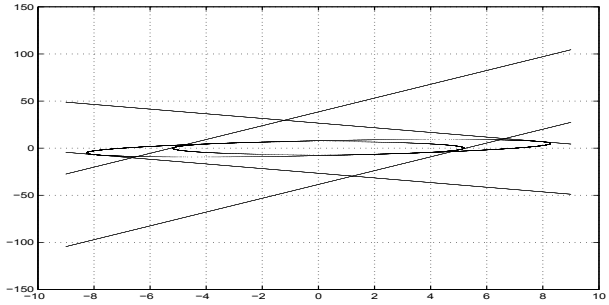


Fig. 6. Inclusion of the ellipsoids inside the polyhedral sets

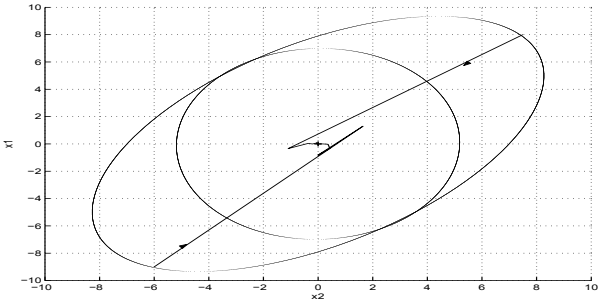


Fig. 7. Motion of the system with controllers obtained with Theorem2.6 and Corollary 3.1

.

Figure 7 shows the level set of stability $\bigcup_{i=1}^{N} \varepsilon(P_i,\ 1)$ using the second method of Theorem 3.2 which is well contained inside the sets of saturations. The use of Theorem 3.3 leads to the following results:

$$K_1 = -0.0752, \quad K_2 = 0.1386;$$

Figure 9 shows the level set of stability $\bigcup_{i=1}^{N} \varepsilon(P_i,\ 1)$ obtained with Theorem 3.3, which is also well contained inside the sets of saturations.

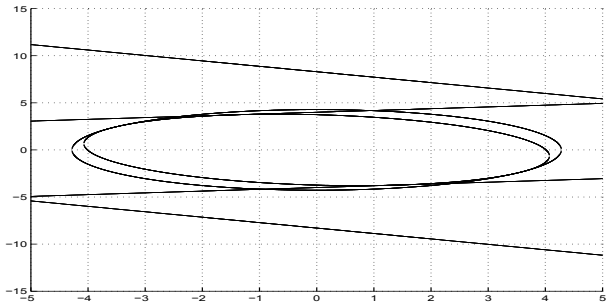


Fig. 8. Inclusion of the ellipsoids inside the polyhedral sets obtained with Theorem 3.2

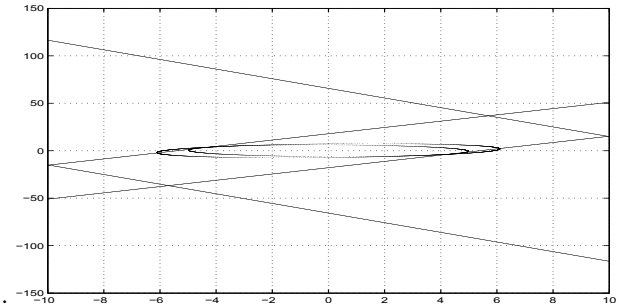


Fig. 9. Inclusion of the ellipsoids inside the polyhedral sets obtained with Theorem 3.3

.

3.3 Synthesis of non saturating controllers

The non saturating controllers who works inside a region of linear behavior can be obtained from the previous results by replacing $D_{si} = \mathbb{I}$ and $D_{si}^{-} = 0$. The following result presents the synthesis of such controllers.

Theorem 3.4. *If there exist symmetric matrices X_i and matrices Y_i such that*

$$\left[\begin{array}{cc} X_i & (A_{i\kappa}X_i + B_{i\kappa}Y_iC_{i\kappa})^T \\ * & X_j \end{array} \right] > 0, \tag{87}$$

$$\left[\begin{array}{cc} 1 & Y_{il}C_{i\kappa} \\ * & X_i \end{array} \right] > 0, \tag{88}$$

$$V_iC_{i\kappa} = C_{i\kappa}X_i, \tag{89}$$

$\forall\ \kappa = 1, \ldots, \mu_i, \quad \forall (i,j) \in \mathcal{I}^2, \forall l \in [1,m]$, *with*, $K_i = Y_iV_i^{-1}, P_i = X_i^{-1}$,
then the uncertain closed-loop switching system (69) is asymptotically stable $\forall\ x_0 \in \Omega$ *and for all switching sequences* $\alpha(t)$.

To illustrate this result, the same system of Example 3.1 is used. Theorem 3.4 leads to the following results:

$$P_1 = \left[\begin{array}{cc} 0.2574 & 0 \\ 0 & 0.2574 \end{array} \right]; P_2 = \left[\begin{array}{cc} 0.2930 & 0.0535 \\ 0.0535 & 0.3376 \end{array} \right];$$

$$K_1 = -0.0902; \quad K_2 = 0.1694.$$

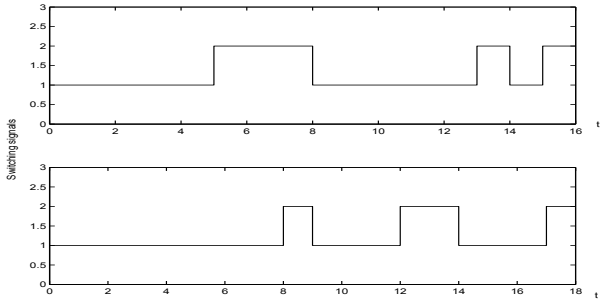


Fig. 10. Switching supervisor signal $\alpha(t)$

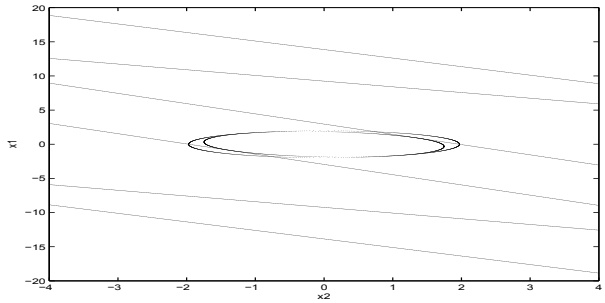


Fig. 11. Inclusion of the ellipsoids obtained with Theorem 3.4

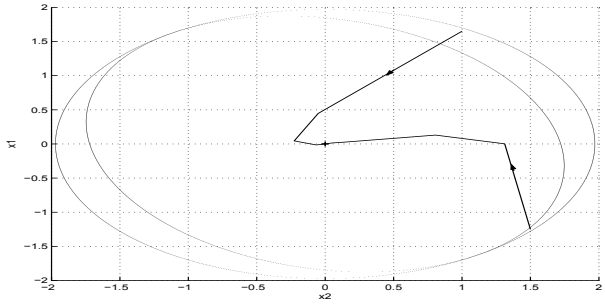


Fig. 12. Motion of the system with controllers obtained with Theorem 3.4

In Figure 5, the evolution of uncertainties is plotted, Figure 10 shows the sequence $\alpha(t)$. The level set $\bigcup_{i=1}^{N} \varepsilon(P_i,\ 1)$ presented in Figure 11, is also well contained inside the regions of linear behavior. In Figure 12, the trajectories of the system are plotted.

Commennt 3.1. *The application of all the proposed results to the same example, shows that the result applied in two steps (Theorem 2.6 and Corollary 3.1) is the least conservative. However, it is worth noting that the result introducing slack variables (Theorem 3.3) is also less conservative even applied in one step. One can expect that this same result applied in two steps can be the less conservative one.*

In this section, two different sufficient conditions of asymptotic stability are obtained for output feedback control of uncertain switching discrete-time linear systems subject to actuator saturations. These conditions allow the synthesis of stabilizing controllers inside a large region of saturation under LMIs formulation. Note that the state feedback control case and the unsaturating controller case can be obtained as particular cases of the study presented in this section. An illustrative example is studied by using the direct resolution of the proposed LMIs. A comparison of the obtained solutions is also given.

4. Stabilization of saturated switching systems with structured uncertainties

The objective of this section is to extend the results of (Benzaouia *et al.,* 2006) to uncertain switching systems subject to actuator saturations by using output feedback control. This technique allows to design stabilizing controllers by output feedback for switching discrete-time systems despite the presence of actuator saturations and uncertainties on the system parameters. The case of state feedback control is derived as a particular case. It is also shown that the results obtained in this section with state feedback control are less conservative than those presented in (Yu *et al.,* 2007) where only the state feedback control case is addressed. The main results of this section are published in (Benzaouia *et al.,* 2009c).

4.1 Problem presentation

Let us consider the linear uncertain discrete-time switching system described by:

$$\begin{cases} x_{t+1} = \mathcal{A}_\alpha(t)x_t + \mathcal{B}_\alpha(t)sat(u_t) \\ y_t = \mathcal{C}_\alpha(t)x_t \end{cases} \tag{90}$$

where $x_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^m$ are the state and the input respectively, $sat(.)$ is the standard saturation, $y_t \in \mathbb{R}^p$ the output. α is a switching rule taking its values in the finite set $I = \{1, ..., N\}$. The saturation function is assumed here to be normalized, *i. e.*, $|sat(u_i| = min(1, |u_i|), i = 1, \dots m$.

The system matrices are assumed to be uncertain and satisfy:

$$[\mathcal{A}_i(t) \quad \mathcal{B}_i(t)] = [A_i \quad B_i] + M_i\Gamma_i\,[N_{1i} \quad N_{2i}] \tag{91}$$

Let the control be obtained by an output feedback control law:

$$u_t = K_\alpha y_k = K_\alpha C_\alpha x_t = F_\alpha x_t$$

The closed-loop system is given by:

$$x_{t+1} = \mathcal{A}_\alpha(t)x_t + \mathcal{B}_\alpha(t)sat(K_\alpha C_\alpha x_t) \tag{92}$$

Defining the indicator function:

$$\xi(t) := [\xi_1(t), ..., \xi_N(t)]^T \tag{93}$$

where $\xi_i(t) = 1$ if the switching system is in mode i and 0 otherwise, yields the following representation for the closed-loop system:

$$x_{t+1} = \sum_{i=1}^{N} \xi_i(t)[\mathcal{A}_i(t)x_t + \mathcal{B}_i(t)sat(K_i C_i x_t)] \tag{94}$$

Assume that there exist N matrices $H_1, \ldots, H_N$ such that $x(t) \in \mathcal{L}(H_i)$. Using the expression in (10) and rewriting System (94) yields that:

$$\begin{aligned} x_{t+1} &= \sum_{s=1}^{\eta} \sum_{i=1}^{N} \xi_i(t)\delta_{is}(t)\mathcal{A}c_{is}(t)x_t; \\ \mathcal{A}c_{is}(t) &:= \mathcal{A}_i(t) + \mathcal{B}_i(t)(D_{is}K_iC_i + D_{is}^{-}H_i), \quad s \in [1, \eta] \end{aligned} \tag{95}$$

4.2 Analysis and synthesis of stabilizability

Consider now the saturated uncertain switching system given by (95). The first result synthesizing stabilizing controllers of the uncertain saturated switching system by output feedback is now presented.

Theorem 4.1. *If there exist symmetric matrices $X_1, \ldots, X_N$, matrices $Y_1, \ldots, Y_N$, $Z_1, \ldots, Z_N$, $V_1, \ldots, V_N$ and a set of real positive scalars λ_{ijs}, such that*

$$\begin{bmatrix} X_i & \Theta_{is}^T & \Phi_{is}^T \\ * & X_j - \lambda_{ijs}M_iM_i^T & 0 \\ * & * & \lambda_{ijs}\mathbb{I} \end{bmatrix} > 0, \tag{96}$$

$$\forall (i,j) \in \mathcal{I} \times \mathcal{I}, \forall s \in [1, \ldots \eta]$$

$$C_iX_i = V_iC_i \tag{97}$$

$$\begin{bmatrix} 1 & Z_{il} \\ * & X_i \end{bmatrix} > 0, \tag{98}$$

$$\forall i \in \mathcal{I}, \forall l \in [1, \ldots m]$$

where $\Theta_{is} = A_iX_i + B_i(D_{is}Y_iC_i + D_{is}^{-}Z_i)$ and $\Phi_{is} = N_{1i}X_i + N_{2i}(D_{is}Y_iC_i + D_{is}^{-}Z_i)$. Then, the uncertain switching system with input saturation in closed-loop (95) with

$$K_i = Y_iV^{-1} \tag{99}$$

$$H_i = Z_iX^{-1} \tag{100}$$

is asymptotically stable $\forall\ x_0 \in \Omega = \bigcup_{i=1}^{N} \varepsilon(X_i^{-1},\ 1)$ and for all switching sequences $\alpha(t)$.

Proof: By using Lemma 2.1, for all $H_i \in \mathbb{R}^{m\times n}$ with $|H_{il}x_t| < 1$, $l \in [1, m]$, there exist $\delta_{i1} \geq 0$,..., $\delta_{i\eta} \geq 0$ such that, $sat(K_iC_ix_t) = \sum_{s=1}^{\eta} \delta_{is}(t)[D_{is}K_iC_i + D_{is}^{-}H_i]x_t$, $\delta_{is}(t) \geq 0$, $\sum_{s=1}^{\eta} \delta_{is}(t) = 1$. System (94) is then rewritten as (95).
Consider the Lyapunov function candidate $V(x) = x_t^T(\sum_{i=1}^{N} \xi_i(t)P_i)x_t$. Computing its rate of increase along the trajectories of system (95) yields:

$$\begin{aligned} \Delta V(x_t) &= x_{t+1}^T(\sum_{j=1}^{N} \xi_j(t+1)P_j)x_{t+1} - x_t^T(\sum_{i=1}^{N} \xi_i(t)P_i)x_t \\ &= \sum_{s=1}^{\eta} \sum_{j=1}^{N} \xi_j(t+1)\delta_{is}x_t^T[\mathcal{A}_i + \mathcal{B}_i(D_{is}F_i + D_{is}^{-}H_i)]^TP_j[\mathcal{A}_i \\ &+ \mathcal{B}_i(D_{is}F_i + D_{is}^{-}H_i)]x_t - \sum_{i=1}^{N} \xi_i(t)x_t^TP_ix_t \end{aligned}$$

Since, $\sum_{s=1}^{\eta}\delta_{is}(t)=\sum_{j=1}^{N}\xi_j(t+1)=\sum_{i=1}^{N}\xi_i(t)=1$, one should obtain

$$\begin{aligned}\Delta V(x_t) &= \sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{s=1}^{\eta}\xi_i(t)\xi_j(t+1)\delta_{is}(t)x_t^T \\ &\quad \Big([\mathcal{A}_i+\mathcal{B}_i(D_{is}F_i+D_{is}^{-}H_i)]^T P_j[\mathcal{A}_i+\mathcal{B}_i(D_{is}F_i+D_{is}^{-}H_i)]-P_i\Big)\,x_t\end{aligned}$$

A sufficient condition to obtain $\Delta V(x_t)<0$ is that:

$$\Big[\mathcal{A}_i+\mathcal{B}_i(D_{is}F_i+D_{is}^{-}H_i)]^T P_j[\mathcal{A}_i+\mathcal{B}_i(D_{is}F_i+D_{is}^{-}H_i)\Big]-P_i=-\Psi_{sij}<0 \tag{101}$$

By applying Schur complement to (101), the following equivalent inequality is obtained:

$$\left[\begin{array}{cc} P_i & [\mathcal{A}_i+\mathcal{B}_i(D_{is}K_iC_i+D_{is}^{-}H_i)]^T \\ * & P_j^{-1}\end{array}\right]>0, \tag{102}$$

Letting $X_i=P_i^{-1}$, $Y_i=K_iV_i$, $C_iX_i=V_iC_i$, $Z_i=H_iX_i$ and multiplying the above inequality on both sides by $diag(X_i,\mathbb{I})$ we get

$$\left[\begin{array}{cc} X_i & [\mathcal{A}_iX_i+\mathcal{B}_i(D_{is}K_iC_i+D_{is}^{-}H_i)X_i]^\top \\ * & X_j\end{array}\right]>0, \tag{103}$$

Taking account of (91), inequality (103) can be developed as follows:

$$-\left[\begin{array}{cc} X_i & [A_iX_i+B_i(D_{is}Y_iC_i+D_{is}^{-}Z_i)]^\top \\ * & X_j\end{array}\right]+\left[\begin{array}{c} [N_{1i}X_i+N_{2i}(D_{is}Y_iC_i+D_{is}^{-}Z_i)]^\top \\ 0\end{array}\right]$$
$$\Gamma_i^\top\left[\begin{array}{cc} 0 & -M_i^T\end{array}\right]+\left[\begin{array}{c} 0 \\ -M_i\end{array}\right]\Gamma_i\left[\begin{array}{cc} [N_{1i}X_i+N_{2i}(D_{is}Y_iC_i+D_{is}^{-}Z_i)] & 0\end{array}\right]<0,$$

by virtue of Lemma2.2, this inequality holds if and only if there exist positive scalars λ_{ijs} such that

$$-\left[\begin{array}{cc} X_i & \Theta_{is}^T \\ * & X_j\end{array}\right]+\lambda_{ijs}\left[\begin{array}{c} 0 \\ -M_i\end{array}\right]\left[\begin{array}{cc} 0 & -M_i^T\end{array}\right]+\frac{1}{\lambda_{ijs}}\left[\begin{array}{cc} \Phi_{is} & 0\end{array}\right]\left[\begin{array}{c} \Phi_{is}^T \\ 0\end{array}\right]<0,$$
$$\forall(i,j)\in\mathcal{I}\times\mathcal{I},\forall s\in[1,\dots\eta].$$

Or in a compact form,

$$\left[\begin{array}{cc} X_i-\frac{1}{\lambda_{ijs}}\Phi_{is}\Phi_{is}^T & \Theta_{is}^T \\ * & X_j-\lambda_{ijs}M_iM_i^T\end{array}\right]>0, \tag{104}$$
$$\forall(i,j)\in\mathcal{I}\times\mathcal{I},\forall s\in[1,\ \eta]$$

where Φ_{is} and Θ_{is} are defined before.
By Schur complement, inequality (104) is equivalent to (96). One can then bound the rate of increase as follows,

$$\Delta V(x_t)\leq-\gamma(\|x_t\|);$$
$$\gamma(\|x_t\|)=min_{ijs}\lambda_{min}(\Psi_{ijs})\|x_t\|^2.$$

Using (Hu *et al.*, 2002), the inclusion condition (29) can also be transformed to the equivalent LMI (98) by virtue of the results of (Boyd *et al.*, 1994). □

To obtain larger ellipsoid domains $\varepsilon(P_i,\ 1)$, we can use a shape reference set $\mathcal{X}_R \subset \mathbb{R}^n$, in terms of a polyhedron or ellipsoid to measure the size of the domain of attraction.

For a set $\mathcal{L} \subset \mathbb{R}^n$ which contains the origin, define $\mu(\mathcal{X}_R, \mathcal{L}) = sup\{\mu > 0,\ \ \mu\mathcal{X}_R \subset \mathcal{L}\}$. Here, we choose $\mathcal{X}_R$ to be a polyhedral defined as $\mathcal{X}_R = co\left\{\omega_1, \omega_2,\ \ldots\ , \omega_q\right\}$, where $\omega_1, \omega_2,\ \ldots\ , \omega_q$ are a prior given points in $\mathbb{R}^n$.

The problem can be formulated as the following constrained optimization problem

$$(Pb.4): \begin{cases} \max_{X_i>0, Y_i, Z_i, \lambda_{ijs}} (\mu_i) \\ s.t.\ \ \mu\mathcal{X}_R \subset \varepsilon(P_i, 1) \\ \quad (96), (98), \\ \quad i = 1, \ldots, N \end{cases}$$

As is explained in (Hu *et al.*, 2001) and (Hu and Lin, 2002), the constraint $\mu\mathcal{X}_R \subset \varepsilon(P_i, 1)$ is satisfied if the following matrix inequalities hold:

$$\begin{bmatrix} \mu_i^{-2} & \omega_l^T \\ \omega_l & X_i \end{bmatrix} \geq 0, \qquad \forall i \in \mathcal{I}, \quad \forall l \in [1,\ q] \tag{105}$$

The problem of enlarging the domain of attraction can be reduced to an LMI optimization problem defined as follows:

$$(Pb.5): \begin{cases} \min_{X_i>0, Y_i, Z_i, \lambda_{ijs}} (\gamma_i) \\ s.t.\ \ (96), (98), (105) \\ \quad i = 1, \ldots, N \end{cases}$$

where $\gamma_i = \mu_i^{-2}$.

Commennt 4.1. *The results of Theorem 4.1 applies directly to switching systems with state feedback control by taking $C_i = \mathbb{I}$. In this case, these results can be compared to the one given in (Yu et al.,2007). The fact that the scalars λ_{ijs} are all kept equal in (Yu et al.,2007), makes the result obviously more conservative. An example will show this conservatism.*

In order to more improve the result of Theorem 4.1 by introducing additional slack variables, the following corollary is presented.

Corollary 4.1. *If there exist symmetric matrices $X_i > 0$, matrices G_i, Y_i, V_i, Z_i and positive scalars λ_{ijs} such that*

$$\begin{bmatrix} G_i^T + G_i - X_i & Y_{is}^T & 0 & \Lambda_{is}^T \\ * & X_j & \lambda_{ijs} M_i & 0 \\ * & * & \lambda_{ijs}\mathbb{I} & 0 \\ * & * & * & \lambda_{ijs}\mathbb{I} \end{bmatrix} > 0, \tag{106}$$

$$\forall (i,j) \in \mathcal{I}^2, \forall s \in [1, \ldots \eta]$$

$$C_i G_i = V_i C_i \tag{107}$$

$$\begin{bmatrix} 1 & Z_{il} \\ * & G_i^T + G_i - X_i \end{bmatrix} > 0, \tag{108}$$

$$\forall i \in \mathcal{I}, \forall s \in [1, \dots \eta], \forall l \in [1, \dots m]$$

where $\Upsilon_{is} = A_i G_i + B_i(D_{is} Y_i C_i + D_{is}^- Z_i)$ *and* $\Lambda_{is} = N_{1i} G_i + N_{2i}(D_{is} Y_i C_i + D_{is}^- Z_i)$. *Then, the uncertain switching system with input saturation in the closed-loop (95) with*

$$K_i = Y_i V^{-1} \tag{109}$$

$$H_i = Z_i G^{-1} \tag{110}$$

is asymptotically stable $\forall\ x_0 \in \Omega = \bigcup_{i=1}^{N} \varepsilon(X_i^{-1},\ 1)$ *and for all switching sequences* $\alpha(t)$.

Proof: It was proven in (Benzaouia *et al.*, 2004) and (Benzaouia *et al.*, 2006) that condition (102) is feasible if and only if there exists non singular matrices G_i such that the following inequality holds:

$$\begin{bmatrix} G_i + G_i^T - X_i & G_i^T[\mathcal{A}_i + \mathcal{B}_i(D_{is} K_i C_i + D_{is}^- H_i)]^T \\ * & X_j \end{bmatrix} > 0, \tag{111}$$

$$\forall (i,j) \in \mathcal{I} \times \mathcal{I}, \forall s \in [1,\ \eta]$$

where $X_i = P_i^{-1}$. The same reasoning is then followed as in the proof of Theorem 4.1 leading to (106). Inequality (108) was also proven in (Benzaouia *et al.*, 2006) by using (Boyd *et al.*, 1994). □

These results can be illustrated with the following example.

Example 4.1. *Consider a SISO saturated switching discrete-time system with two modes given by the following matrices:*

$$\begin{aligned} A_1 &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, B_1 = \begin{bmatrix} 10 \\ 5 \end{bmatrix}, M_1 = 0.1\mathbb{I}, N_{11} = N_{12} = 0.01\mathbb{I}, \\ A_2 &= \begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix}, B_2 = \begin{bmatrix} 0.5 \\ -2 \end{bmatrix}, M_2 = 0.1\mathbb{I}, N_{21} = N_{22} = 0.01\mathbb{I}, \end{aligned}$$

By solving the optimization problem (Pb.5) for the above system, we can obtain the following results:

$$P_1 = 10E-03 \begin{bmatrix} 4.3324 & 1.2516 \\ 1.2516 & 4.3324 \end{bmatrix}; P_2 = 10E-03 \begin{bmatrix} 4.3988 & 2.0934 \\ 2.0934 & 6.1433 \end{bmatrix},$$

$$\begin{aligned} H_1 &= [-0.0261536 \quad -0.0653823]; H_2 = [-0.0000192 \quad 0.0717335] \\ K_1 &= -0.1000089; K_2 = 0.1256683 \end{aligned}$$

The corresponding figures are given by Figure13 and Figure 14. By applying Corollary 4.1, the following results are obtained:

$$P_1 = 10E-03 \begin{bmatrix} 1.0047 & 0.1917 \\ 0.1917 & 2.0626 \end{bmatrix}; P_2 = 10E-04 \begin{bmatrix} 7.380 & 1.823 \\ 1.823 & 23.530 \end{bmatrix},$$

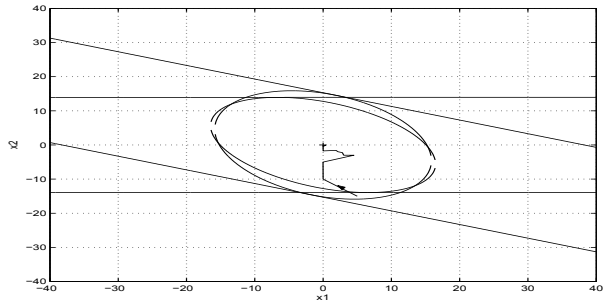


Fig. 13. Inclusion of the ellipsoids inside the polyhedral sets using Theorem 4.1

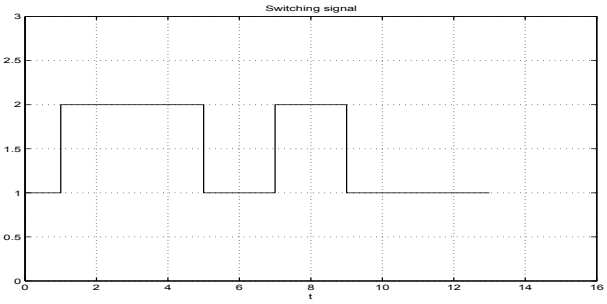


Fig. 14. The switching sequences

$$\begin{aligned} H_1 &= [-0.0179753 \quad -0.0405034]; H_2 = [-0.0000072 \quad 0.0480397] \\ K_1 &= -0.0597163; K_2 = 0.0365389. \end{aligned}$$

The corresponding figures are given by Figure 15 and Figure 16. In order to compare the present results with (Yu et al., 2007) for state feedback control, the same example is considered. The results of Theorem 4.1 give:

$$P_1 = 10E-08 \left[\begin{array}{cc} 8.656 & 60 \\ 60 & 6900 \end{array} \right] \qquad P_2 = 10E-08 \left[\begin{array}{cc} 1.691 & 7.501 \\ 7.501 & 6750 \end{array} \right],$$

$$\begin{aligned} H_1 &= [-0.0002470 \quad -0.0060812]; H_2 = [-0.0000081 \quad 0.0081436] \\ F_1 &= [-0.0003254 \quad -0.0099873]; F_2 = [-0.0068339 \quad 0.6156659] \end{aligned}$$

while the results of (Yu et al., 2007) give:

$$P_1 = 10E-07 \left[\begin{array}{cc} 2 & 8 \\ 8 & 773 \end{array} \right]; P_2 = 10E-08 \left[\begin{array}{cc} 4.894 & 20 \\ 20 & 7630 \end{array} \right],$$

$$\begin{aligned} H_1 &= [-0.0003203 \quad -0.0066782]; H_2 = [-0.0000016 \quad 0.0086701] \\ F_1 &= [-0.0050262 \quad -0.3821976]; F_2 = [-0.0112893 \quad 0.3695061] \end{aligned}$$

The corresponding level sets are depicted in Figure 17 and Figure 18 where the conservatism of the results of (Yu et al., 2007) is obvious.

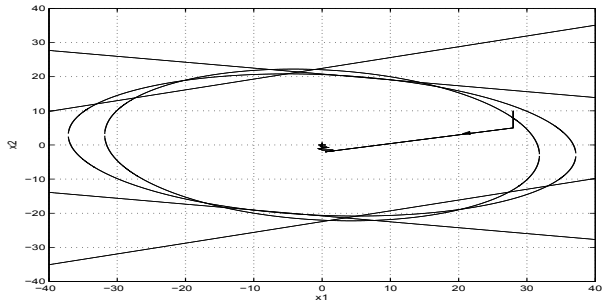


Fig. 15. Inclusion of the ellipsoids inside the polyhedral sets using Theorem 4.1

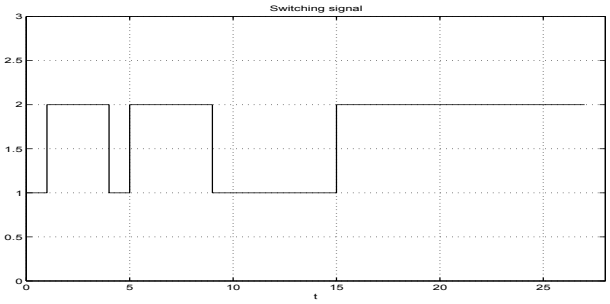


Fig. 16. The switching sequences

This section studied uncertain switching systems with output feedback control which extends the results of (Yu *et al.*, 2007) given with state feedback control. In order to compare these two results, a numerical study using only the particular case of the present work, as mentioned by the Comment 4.1, is also presented. The obtained improvements with the methods presented in this work are shown in Figure 17 and Figure 18. A numerical example is used to illustrate all these techniques.

5. CONCLUSION

In this chapter, two main different sufficient conditions of asymptotic stability are obtained for switching discrete-time linear systems subject to actuator saturations for each case: state feedback and output feedback control. The first allows the synthesis of stabilizing controllers inside a large region of linear behavior while the second applies the idea of Lemma (Hu *et al.*, 2002) which rewrites the saturation function under a combination of 2^m elements to obtain stabilizing controllers tolerating saturations to take effect. A particular attention is given to the output feedback case which has additive complexity due to the output equation. In this sense, three different LMIs are presented for this case. The main results of this work are given under LMIs formulation leading to the design of the stabilizing state feedback and output feedback controllers for the system. Even the dynamical system is a switching system, it is shown that the set Ω formed by the union of all the ellipsoid level sets associated to each subsystem, constitutes a set of asymptotic stability. The first time that this important result is established for saturated switching systems is in (Benzaouia *et al.*, 2006). Two illustrative examples are studied by using the solution of the proposed LMIs. A comparison of the obtained

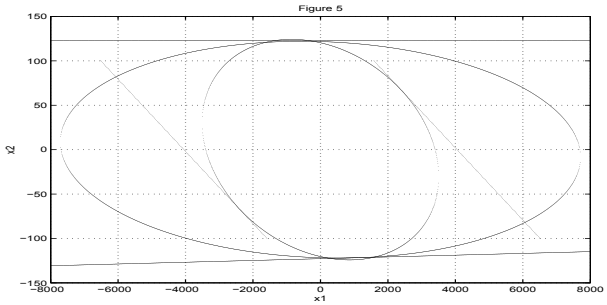


Fig. 17. Inclusion of the ellipsoids inside the polyhedral sets using Theorem 4.1

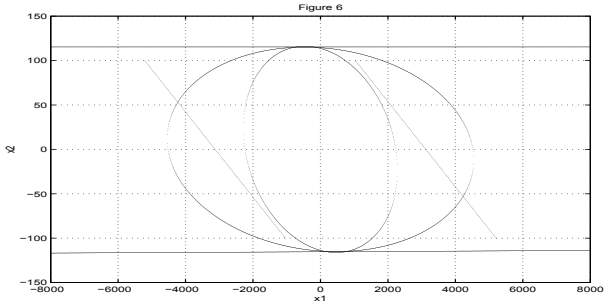


Fig. 18. Inclusion of the ellipsoids inside the polyhedral sets using (Yu *et al.*, 2007)

solutions is also given.
Further, sufficient conditions of stabilization of switching linear discrete-time systems with polytopic and structured uncertainties are also obtained. These conditions are given under LMIs form. Both the cases of feedback control and output control are studied for polytopic uncertainties. However, for structured uncertainties, the output feedback control is presented extending the results of (Yu *et al.*, 2007) given with state feedback control. A comparison study is given with a numerical particular case. The obtained improvements with our method are also shown. A numerical example is used to illustrate all these techniques. As a perspective, two new works developed for switching systems without saturation, the first concerns positive switching systems (Benzaouia and Tadeo, 2008) while the second concerns the output feedback problem (Bara and Boutayeb, 2006) can be used with saturated controls.

6. REFERENCES

G. I. Bara and M. Boutayeb, "Switched output feed back stabilization of discrete-time switched systems". 45^{th} *Conference on Decision and Control, December 13-15*, San Diego, pp. 2667-2672, 2006.

A. Benzaouia, C. Burgat, "Regulator problem for linear discrete-time systems with non-symmetrical constrained control". *Int. J. Control*. Vol. 48, N° 6, pp. 2441-2451, 1988.

A. Benzaouia, A. Hmamed, "Regulator Problem for Linear Continuous Systems with Non-symmetrical Constrained Control". *IEEE Trans. Aut. Control*, Vol. 38, N° 10, pp. 1556-1560, 1993.

A. Benzaouia and A. Baddou, "Piecwise linear constrained control for continuous time systems". *IEEE Trans. Aut. Control*, Vol. 44, N° 7 pp. 1477-1481, 1999.

A. Benzaouia, A. Baddou and S. Elfaiz, "Piecewise linear constrained control for continuous-time systems: An homothetic expansion method of the initial domain. Journal of Dynamical and Control Systems. Vol. 12, N°. 2 (April), pp. 277-287, 2006.

A. Benzaouia, L. Saydy and O. Akhrif, "Stability and control synthesis of switched systems subject to actuator saturation". *American Control Conference, June 30- July 2*, Boston, 2004.

A. Benzaouia, O. Akhrif and L. Saydy, "Stability and control synthesis of switched systems subject to actuator saturation by output feedback". 45^{th} *Conference on Decision and Control, December 13-15*, San Diego, 2006.

A. Benzaouia, F. Tadeo and F. Mesquine, "The Regulator Problem for Linear Systems with Saturations on the Control and its Increments or Rate: An LMI approach". *IEEE Transactions on Circuit and Systems Part I*, Vol. 53, N°. 12, pp. 2681-2691, 2006.

A. Benzaouia, E. DeSantis, P. Caravani and N. Daraoui, "Constrained Control of Switching Systems: A Positive Invariance Approach". Int. J. of Control, Vol. 80, Issue 9, pp. 1379-1387, 2007.

A. Benzaouia and F. Tadeo. "Output feedback stabilization of positive switching linear discrete-time systems". 16^{th} *Mediterranean Conference, Ajaccio, France June 25-27*, 2008.

A. Benzaouia, O. Akhrif and L. Saydy. "Stabilitzation and Control Synthesis of Switching Systems Subject to Actuator Saturation". Int. J. Systems Sciences. To appear 2009.

A. Benzaouia, O. Benmesaouda and Y. Shi " Output feedback Stabilization of uncertain saturated discrete-time switching systems". IJICIC. Vol. 5, N°. 6, pp. 1735-1745, 2009.

A. Benzaouia, O. Benmesaouda and F. Tadeo. "Stabilization of uncertain saturated discrete-time switching systems". Int. J. Control Aut. Sys. (IJCAS). Vol. 7, N°. 5, pp. 835-840, 2009.

F. Blanchini, "Set invariance in control - a survey". *Automatica*, Vol. 35, N°. 11, pp. 1747-1768, 1999.

F. Blanchini and C. Savorgnan, "Stabilizability of switched linear systems does not imply the existence of convex Lyapunov functions". 45^{th} *Conference on Decision and Control, December 13-15*, San Diego, pp. 119-124, 2006.

F. Blanchini, S. Miani and F. Mesquine, "A Separation Principle for Linear Switching Systems and Parametrization of All Stabilizing Controllers, "IEEE Trans. Aut. Control", Vol. 54, No. 2, pp. 279-292, 2009.

E. L. Boukas, A. Benzaouia. "Stability of discrete-time linear systems with Markovian jumping parameters and constrained Control". *IEEE Trans. Aut. Control*. Vol. 47, N°. 3, pp. 516-520, 2002.

S. P. Boyd, EL Ghaoui, E. Feron, and V. Balakrishnan. "Linear Matrix Inequalities in System and Control Theory". SIAM, Philadelphia, PA, 1994.

M. S. Branicky, "Multiple Lyapunov functions and other analysis tools for switched and hybrid systems". IEEE Automat. Contr., Vol. 43, pp. 475-482, 1998.

E. F. Camacho and C.Bordons, "Model Predictive Control"', Springer-Verlag, London, 2004.

M. Chadli, D. Maquin and J. Ragot. "An LMI formulation for output feedback stabilization in multiple model approach". In Proc. of the 41 *th* CDC, Las Vegas, Nevada, 2002.

J. Daafouz and J. Bernussou. "Parameter dependent Lyapunov functions for discrete-time systems with time varying parametric uncertainties", *Systems and Control Letters*, Vol. 43, No. 5, pp. 355-359, 2001.

J. Daafouz, P. Riedinger and C. Iung. "Static output feedback control for switched systems". *Procceding of the* 40*th IEEE Conference on Decision and Control, Orlando, USA*, 2001.

J. Daafouz, P. Riedinger and C. Iung. "Stability analysis and control synthesis for switched systems: a switched Lyapunov function approach". *IEEE Trans. Aut. Control*, Vol. 47, N°. 11, pp. 1883-1887, 2002.

L. El Ghaoui, F. Oustry, and M. AitRami "A Cone Complementarity Linearization Algorithm for Static Output-Feedback and Related Problems". IEEE Trans. Aut. Control, Vol.42, N°8, pp.1171 − 1176, 1997.

L. Hetel, J. Daafouz, and C. Iung. "Stabilization of Arbitrary Switched Linear Systems With Unknown Time-Varying Delays". *IEEE Trans. Aut. Control*, Vol. 51, N°. 10, pp. 1668-1674, 2006.

G. Ferrai-Trecate, F. A. Cuzzola, D. Mignone and M. Morari. "Analysis and control with performanc of piecewise affine and hybrid systems". *Procceding of the American Control Conference, Arlington, USA*, 2001.

P. Gutman and P. Hagandar. "A new design of constrained controllers for linear systems," *IEEE Trans. Aut. Cont.*, $Vol.AC - 30, pp.22 - 33$, 1985.

T. Hu, Z. Lin and B. M. Chen, "An analysis and design method for linear systems subject to actuator saturation and disturbance". *Automatica*, Vol. 38, pp. 351-359, 2002.

T. Hu, Z. Lin, "Control Systems with Actuator Saturation: Analysis and Design", BirkhVauser, Boston, 2001.

T. Hu, L. Ma and Z. Lin. "On several composite quadratic Lyapunov functions for switched systems". *Procceding of the* 45^th *IEEE Conference on Decision and Control, San Diego, USA*, pp. 113-118, 2006.

D. Liberzon, "Switching in systems and control". *Springer*, 2003.

J. Lygeros, C. Tomlin, and S. Sastry. "Controllers for reachability specifications for hybrid systems", Automatica, vol. 35, 1999.

D. Mignone, G. Ferrari-Trecate,and M. Morari, "Stability and stabilization of Piecwise affine and hybrid systems: an LMI approach". *Procceding of the 39th IEEE Conference on Decision and Control*, Sydney, Australia, 2000.

E. F. Mulder, M. V. Kothare and and M. Morari, "Multivariable anti-windup controller synthesis using linear matrix inequalities". *Automatica*, Vol. 37, No.9, pp. 1407-1416, 2001.

R. N. Shorten, and Narendra K.S, a) "On the existence of a commun Lyapunov function for linear stable switching systems". *Proc. 10th, Yale Workshop on Adaptive and Learning Systems*, pp.130-140, 1998. b) "A sufficient condition for the existence of a commun Lyapunov function for two second-orderliinear systems". *Proc, 36th Conf. Decision and Control*, pp. 3521-3522,1997.

D. Xie, L. Wang, F. Hao, G. Xie, "Robust Stability Analysis and Control Synthesis for Discret-time Uncertain switched Systems". Conference on Decision and control. Hawaii, USA, 2003.

J. Yu, G. Xie and L. Wang, "Robust Stabilization of discrete-time switched uncertain systems subject to actuator saturation". American Control Conference, New York, July 11-13, 2007.

2

Robust Adaptive Control of Switched Systems

Khalid El Rifai and Kamal Youcef-Toumi
Department of Mechanical Engineering
Massachusetts Institute of Technology
77 Massachusetts Ave. Room 3-350
Cambridge, MA 02139, USA

Abstract
In this chapter, a methodology for robust adaptive control design for a class of switched nonlinear systems is developed. Under extensions of typical adaptive control assumptions, a leakage-type adaptive control scheme guarantees stability for systems with bounded disturbances and parameters without requiring a priori knowledge on such parameters or disturbances. The problem reduces to an analysis of an exponentially stable and input-to-state stable (ISS) system driven by piecewise continuous and impulsive inputs due to plant parameter switching and variation. As a result, a separation between robust stability and robust performance and clear guidelines for performance optimization via ISS bounds are obtained. The results are demonstrated through example simulations, which follow the developed theory and demonstrate superior robustness of stability and performance relative to non-adaptive and other adaptive methods such as projection and deadzone adaptive controllers.

1. Introduction

Switched and hybrid systems have been gaining considerable interest in both research and industrial control communities. This is motivated by the need for systematic and formal methods to control such systems. These issues arise in systems with discrete changes in energy exchange elements due to intermittent interaction with other systems or with an environment or due to the nature of their constitutive relations. This is common in robotic and mechatronic systems with contact and impact effects, fluidic systems with valves or phase changes, and electrical circuits with switches.
Despite numerous interesting publications on hybrid systems, there is a lack of constructive methods for control of a nontrivial class of switched systems with a priori stability and performance guarantees due to the difficulty of this problem. In terms of stability and response of switched systems, several results have been obtained in recent years, see (10; 2; 25) and references therein. In this context, sufficient conditions for stability such as common Lyapunov functions and average dwell time (10) are the most commonly studied approaches. A corresponding control design requires switching controller gains such that all subsystems are made stable and such that a common Lyapunov function condition is satisfied, which for LTI systems requires system matrices to commute or be symmetric, see (17; 18) for more explicit results. In order to verify that such a condition is met, the system is partitioned into known subsystems and a set of linear matrix inequalities, of increasing order with the number of subsystems, is solved if a solution is feasible. The other class of results requires that

all subsystems are stable (or with some known briefly visited unstable modes) and switching is slow enough on average, *average dwell time condition* (10). The corresponding controller design requires gains to be adjusted to guarantee the stability of each frozen configuration and knowledge of worst case decay rate among subsystems and condition number of Lyapunov matrices in order to compute the maximum admissible switching speed. If plant switching exceeds this switching speed then stability can no longer be guaranteed. Analogous analysis results have been extended for systems with disturbances (22) and with some uncertainties (23) as well as related work for linear-parameter varying (LPV) systems in (20; 12). Thus, there is a need for more explicit methods that can be constructively used to *design* controllers for stable switched systems independent of the success of heuristics or feasibility of complex computational methods.

Adaptive control is another popular approach to deal with system uncertainty. The problem with conventional adaptive controllers is that the transient performance is not characterized and stability with respect to bounded parameter variations or disturbances is not guaranteed. Robust adaptive controllers, (6), developed to address the presence of disturbances and non-parametric uncertainties, are typically based on projection, switching-sigma or deadzone adaptation laws that require a priori known bounds on parameters, and in some cases disturbances as well, in order to ensure state boundedness. Extensions to some classes of time varying systems have been developed in (13; 14; 15; 24). However, the results are restricted to smoothly varying parameters with known bounds and typically require additional restrictive conditions such as slowly varying unknown parameters (24) or constant and known input vector parameters (14), in order to ensure state boundedness. In this case, such a conclusion is of very little practical importance if the error can not be reduced to an acceptable level by increasing the adaptation or feedback gains or using a better nominal estimate of the plant parameters. Furthermore, performance with respect to rejection of disturbances as well as the transient response remain primarily unknown.

However, a leakage-type modification as will be shown in this chapter, achieves internal exponential stability and input-to-state stability (ISS), for the class of systems under consideration, without need for persistence of excitation as required in (6). In this regard, projection and switching-sigma modifications have been favored over fixed-sigma modifications, (6) due to its inability to achieve zero steady-state tracking when parameters are constant and disturbances vanish. However, this is a situation of no interest to this paper since the focus is on time varying switching systems. The developed control methodology, which is a generalization of fixed-sigma modification, yields strong robustness to time varying and switching parameters without requiring a priori known bounds on such parameters, as typically needed in projection and switching-sigma modifications.

In this chapter, the development and formulation of an adaptive control methodology for a class of switched nonlinear systems is presented. Under extensions of typical adaptive control assumptions, a leakage-type adaptive control scheme is developed for systems with piecewise differentiable bounded parameters and piecewise continuous bounded disturbances without requiring a priori knowledge on such parameters or disturbances. This yields a separation between robust stability and robust performance and clear guidelines for performance optimization via ISS bounds.

The remainder of the chapter is organized as follows. Section 2 presents the basic adaptive controller methodology. Analysis of the performance of the control system along with design guidelines is discussed in Section 3. Section 4 gives an example simulation demonstrating the key characteristics of the control system as well as comparing it with other non-adaptive

and adaptive techniques such as projection and dead-zone. Conclusions are given in Section 5. In this chapter, $\overline{\lambda}(.)$ and $\underline{\lambda}(.)$ denote the maximal and minimal eigenvalues of a symmetric matrix, $\|.\|$ the euclidian norm, and $diag(.,.,\dots)$ denotes a block diagonal matrix.

2. Methodology

2.1 Parameterized Switched Systems

A hybrid switched system is a system that switches between different vector fields in a differential equation (or a difference equation) each active during a period of time. In this chapter we consider feedback control of continuous-time switched time varying systems described by:

$$\begin{aligned} \dot{x}(t) &= f_i(x,t,u,d),\ t_{i-1} \le t < t_i \\ y(t) &= h_i(x,t),\ t_{i-1} \le t < t_i \\ i(t)^+ &= g(i(t),x,t) \end{aligned} \tag{1}$$

where x is the continuous state, d is for disturbances, u is the control input and y is measured output. Furthermore, $i(t) \in \{1,2,3\dots\}$ is a piecewise constant signal with i denoting the i^{th} switched subsystem active during a time interval $[t_{i-1},t_i)$, where t_i is the i^{th} switching time. The signal $i(t)$, usually referred to as the switching function, is the discrete state of this hybrid system. The discrete state is governed by the discrete dynamics of $g(i(t),x,t)$, which sees the continuous state x as an input. This means switching may be triggered by a time event or a state event, e.g. x reaching certain threshold values, or even memory, i.e, past values for $i(t)$. on state only implicitly with enforced

In this chapter, we view a switching system as one parameterized by a time varying vector of parameters, which is piecewise differentiable, see Equation (2). This is a reasonable representation since it captures many physical systems that undergo switching dynamics, thus we will focus on such systems described by:

$$\begin{aligned} \dot{x} &= f(x,a,u,d) \\ y &= h(x,a) \\ a(t) &= a_i(t),\ t_{i-1} \le t < t_i,\ i = 1,2,\dots \\ i(t)^+ &= g(i(t),x,t) \end{aligned} \tag{2}$$

Therefore, we embed the switching behavior in the piecewise changes in $a(t)$, which again may be triggered by state or time driven events. $a_i(t) \in C^1$, i.e., at least one time continuously differentiable. This means $a(t)$ is piecewise continuous, with a well defined bounded derivative everywhere except at points t_i where $\dot{a} = da/dt$ consists of dirac-delta functions. Also the points of discontinuity of a, which are distinct and form an infinitely countable set, are separated by a nonzero dwell time, i.e., there are no Zeno phenomena (11; 21). This is a reasonable assumption since this is how most physical systems behave. The main assumptions on the class of systems under consideration are formally stated below:

Assumption 1
For a switched system given by Equation (2) the set of switches associated with a switching sequence $\{(t_i,a_i)\}$ *is infinitely countable and* $\exists$ *a scalar* $\mu > 0$ *such that* $t_i - t_{i-1} \ge \mu\ \forall i$.

Assumption 2 $d \in \mathbb{R}^k$ *is uniformly bounded and piecewise continuous.*

Assumption 3 *$a \in \mathcal{S}_a$ is uniformly bounded and piecewise differentiable, where the set $\mathcal{S}_a$ is an admissible, but not necessarily known, set of parameters.*

Note that by allowing piecewise changes in a the parametrization allows structural changes in the system if we overparametrize such that all possible structural terms are included. Then some parameters may switch to or from the value of zero as structural changes take place in the system.

2.2 Robust Adaptive Control

In this section, we discuss the basic methodology based on observation of the general structure of the adaptive control problem. In standard adaptive control for linearly-parameterized systems we usually have control and adaptation laws of the form:

$$\begin{aligned} u &= g(x_m, \hat{a}, \dot{\hat{a}}, y_r, t) \\ \dot{\hat{a}} &= f_a(x_m, \hat{a}, y_r, t) \end{aligned} \tag{3}$$

where u is the control signal, $\hat{a}$ is an estimate of plant parameter vector $a \in S_a$, where S_a is an admissible set of parameters, x_m is measured state variables, and y_r is a desired reference trajectory to be followed. This yields the following closed loop error dynamics :

$$\begin{aligned} \dot{e}_c &= f_e(e_c, \tilde{a}, t) + d(t) \\ \dot{\tilde{a}} &= f_a(e_c, \hat{a}, t) - \dot{a} \end{aligned} \tag{4}$$

where e_c represents a generalized tracking error vector, which includes state estimation error in general output feedback problems and can depend nonlinearly on the plant states as in backstepping designs, $\tilde{a} = \hat{a} - a$ is parameter estimation error, and d is the disturbance.
In standard adaptive control we typically design the control and adaptation laws, Equation (3), such that $\forall\, a \in S_a$ we have:

$$e_c^T P f_e + \tilde{a}^T \Gamma(t)^{-1} f_a \quad \leq \quad -e_c^T C e_c \tag{5}$$

where matrices $P > 0$ and $C > 0$ are chosen depending on the particular algorithm, e.g. choice of reference model and the diagonal matrix $\Gamma(t)^{-1} = diag(\Gamma_o^{-1}, \gamma_\rho^{-1}|b(t)|) > 0$ is an equivalent generalized adaptation gain matrix, where diagonal matrix $\Gamma_o > 0$ and scalar $\gamma_\rho > 0$ are the actual adaptation gains used in the adaptation laws. Whereas, $b(t)$ is a scalar plant parameter, usually the high frequency gain, which appears in Γ in some adaptive designs. The following additional assumption is made for $b(t)$:

Assumption 4 *$b(t)$ is an unknown scalar function such that $b(t) \neq 0\ \forall t$, and sign of $b(t)$ is known and constant.*

This is sufficient to stabilize the system with constant parameters and no disturbances. However, since the error dynamics is not ISS stable, stability is no longer guaranteed in the presence of bounded inputs such as d and $\dot{a}$. In order to deal with time varying and switching dynamics, a modification to the adaptation law will be pursued.
Now consider the following modified adaptation law:

$$\dot{\hat{a}} = f_a(e_c, \hat{a}, t) - L(\hat{a} - a^*) \tag{6}$$

with the diagonal matrix $L = diag(L_o, L_\rho) > 0$ and $a^*(t)$ is an arbitrarily chosen piecewise continuous bounded vector, which is an additional estimate of the plant parameter vector. Then the same system in Equation (4) with the modified adaptation law becomes:

$$\begin{aligned} \dot{e}_c &= f_e(e_c, \tilde{a}, t) + d(t) \\ \dot{\tilde{a}} &= f_a(e_c, \hat{a}, t) - L\tilde{a} + L(a^* - a) - \dot{a} \end{aligned} \tag{7}$$

The modified adaptation law shown above is similar to leakage adaptive laws (6), which have been used to improve robustness with respect to unstructured uncertainties. The leakage adaptation law, also known as fixed-sigma, uses $L_o = \sigma\Gamma_o$, where $\sigma > 0$ is a scalar and the vector $a^*(t)$ above is usually not included or is a constant. In fact, the key contribution from the generalization presented here is not in the algebraic difference relative to leakage adaptive laws (6) but rather in how the algorithm is utilized and proven to achieve new properties for control of rapidly varying and switching systems. In particular, internal exponential and ISS stability of the closed loop system using this leakage-type adaptive controller, without need for persistence of excitation as required in (6), is shown and used to guarantee stability of the state $x_c = [e_c^T, \tilde{a}^T]^T$, see Theorem 1 below.

Theorem 1 *If there exits matrices $P, \Gamma_o, \gamma_\rho, C > 0$ such that (5) is satisfied for $\dot{a} = d = 0$ with $\Gamma(t)^{-1} = diag(\Gamma_o^{-1}, \gamma_\rho^{-1}|b(t)|) > 0$ and Assumption 2.4 is satisfied then the system given by Equation (7) with $d, \dot{a} \neq 0$ and diagonal $L > 0$ is :*
(i) Uniformly internally exponentially stable and ISS stable.
(ii) If Assumptions (2.1-2.3) are satisfied and $a^*(t)$ is chosen as a piecewise continuous bounded vector then state $x_c = [e_c^T, \tilde{a}^T]^T$ is bounded with

$$\|e_c(t)\| \leq c_1 \|x_c(t_o)\| e^{-\alpha(t-t_o)} + c_2 \int_{t_o}^{t} e^{\alpha(\tau - t)} \|v(\tau)\| d\tau$$

where c_1, c_2 are constants, $\alpha = \bar{\lambda}(diag(P^{-1}C, L))$, and $v = [P^{1/2}d, \Gamma^{-1/2}(L(a^* - a) - \dot{a})]^T$.

The proof of this result is found in Appendix A.

2.3 Remarks

This section presents some remarks summarizing the implications of this result.

- The effect of plant variation and uncertainty is reduced to *inputs* $L(a^* - a)$ and $\dot{a}$ acting on this ISS closed loop system. This, in turn, provides a separation between the robust stability and robust performance control problems.
- The modified adaptation law is a slightly more general version of the leakage modification, also known as fixed-sigma, (6), where $L = \sigma\Gamma$, where $\sigma > 0$ is a scalar and the vector $a^*(t)$ above is usually not included or is a constant. This is a robust adaptive control method that has been less popular than projection and switching-sigma modifications due to its inability to achieve zero steady-state tracking when parameters are constant and disturbances vanish. However, this approach yields stronger stability and performance robustness for time varying switching systems for which the constant parameter case is irrelevant.
- Plant parameter switching no longer affects internal dynamics and stability but enters as a step change in input $L(a^* - a)$ and an impulse in input $\dot{a}$ at the switching instant.

- Controller switching of a^* does not affect internal dynamics but enters as a step change in input $L(a^* - a)$, which is a very powerful feature that can be used to utilize available information about the system.
- Allowed arbitrary time variation and switching in the parameter vector a are for a plant within the admissible set of parameters S_a. This set has not been defined here and will be defined later via design assumptions for the classes of systems of interest.
- The authors believe that the use of this robust adaptive controller is useful for switched systems even in the switched linear uncertainty free plant case, where stability with switched linear feedback is difficult to guarantee based on currently available tools (switching between stable LTI closed loop subsystems does not preserve stability). In this case, knowledge of the switching plant parameter vector $a(t)$ can be used in $a*(t)$.

3. Performance of the Control System

In this section, the tracking performance of the obtained control system is discussed.

3.1 Dynamic Response

Exponential stability allows for shaping the transient response, e.g. settling time, and frequency response of the system to low/high frequency dynamics and inputs by adjusting the decay rate α, see Theorem 1. This is to be done independent of the parametric uncertainty $a^* - a$, which is contrasted to LTI feedback where closed loop poles change with parametric uncertainty. Thus the response to step and impulse inputs is as we expect for such an exponentially stable system. However, in this case such inputs will not arise from only disturbances but also from parameters and their variation. In particular, switches in parameters $a(t)$ yields step changes in a and impulses in $\dot{a}(t)$. Furthermore, the system display the frequency response characteristics such as in-bandwidth input, disturbances and parametric uncertainty and variations, rejection and more importantly attenuation of high frequency inputs due to roll-off.

3.2 Improving Tracking Error

Since stability and dynamic response of the system to different inputs and uncertainties have been established independent of uncertainty, we are now left with optimizing the control parameters and gains a^*, L, Γ, P, and C for minimal tracking error. Different methods for improving tracking error are described below with reference to the bound in Theorem 1:

1. *Increasing the system input-output gain* $\alpha = \underline{\lambda}(diag(P^{-1}C, L))$, which as discussed earlier, acts on the overall input uncertainty v. This attenuation, however, increases the system bandwidth, which suggests its use primarily for low/high bandwidth disturbances along the line of frequency response analysis of last section.
2. *Increasing adaptation gain* Γ, which has the effect of attenuating parametric uncertainty and variation independent of system bandwidth (Recall that α is independent of Γ from Theorem 1). This is the case since the size of the input v is reduced by reducing the component $\Gamma^{-1/2}(L(a^* - a) - \dot{a})$. Note that a very large Γ has the effect of amplifying measurement noise, which can be seen from the adaptation law.
3. *Using a small gain* $\Gamma^{-1/2}L$, which is an agreement with increasing adaptation gain matrix Γ mentioned above. However, this differs by the fact that this can be also achieved by simply reducing the size of L. Furthermore, using $\Gamma^{-1/2}L$ is effective mainly for

parametric uncertainty since the input v contains $\Gamma^{-1/2}(L(a^* - a) - \dot{a})$, which suggests a small $\Gamma^{-1/2}L$ does not necessarily attenuate $\dot{a}$ unless $\Gamma^{-1/2}$ is also small. This is the case since this condition implies having *approximate integral action* in the adaptation law of Equation (7), i.e., approaching integral action in the standard gradient adaptation law.

4. *Adjusting and updating parameter estimate* a^* , which can be any piecewise continuous bounded function. This allows for reducing the effect of parametric uncertainty through reducing size of input $a^* - a$ independent of system bandwidth and control gains. In this regard, many of the useful and interesting ideas to monitor, select, and switch between different candidate controllers via multiple models such as those in (1; 16; 7; 26) can be used with switching between a_i^* values playing the role of the i^{th} candidate controller. The difference is that this is to be done without frozen-time instability or switched system instability concerns (verifying dwell time or common Lyapunov function conditions) as $a^*(t)$ is just an input to the closed loop system. Similarly, gain scheduling and Linear Parameter Varying (LPV) control (12; 20) can be applied with a^* playing the role of the scheduled parameter vector to be varied, again with no concerns with instability and transient behavior since $a^* - a$ enter as an input to the system.

3.3 Remarks

- Exponential stability allows for shaping the transient response, e.g. settling time, and frequency response of the system to low/high frequency dynamics and inputs by adjusting the decay rate α, see Theorem 1. This is to be done independent of the parametric uncertainty $a^* - a$, which is contrasted to LTI feedback where closed loop poles change with parametric uncertainty.
- The attenuation of uncertainty by high input-output system gain in this scheme differs from robust control by the fact that ISS stability, the pre-requisite to such attenuation, is never lost due to large parametric uncertainty $a^* - a$. This is the case since it no longer enters as a function of the plant's state but rather as an input $L(a^* - a)$.
- In switching between different a^* values many of the useful and interesting ideas to monitor, select, and switch between different candidate controllers via multiple models such as those in (1; 16) can be used with a_i^* values playing the role of the i^{th} candidate controller. The difference is that this is to be done without frozen-time instability or switched system instability concerns (verifying dwell time or common Lyapunov function conditions) as a^* is just an input to the closed loop system. Similarly, gain scheduling and Linear Parameter Varying (LPV) control (12; 20) can be applied with a^* playing the role of the scheduled parameter vector to be varied, again with no concerns with instability and transient behavior since $a^* - a$ enter as an input to the system.

4. Example Simulation

Consider the following unstable 2^{nd} order plant of relative degree 1 with a 2-mode periodic switching:

$$\begin{aligned} \dot{x}_1 &= a_1 x_1^3 + x_2 + (1 + x_1^2) b_1 u + d \\ \dot{x}_2 &= a_2 x_1 + (1 + x_1^2) b_2 u \\ y &= x_1 + n \end{aligned}$$

where u, d, and n are control signal, disturbance, and measurement noise respectively. Whereas, the plant parameters are given by:

$$\begin{aligned} a_1 &= 3 + 30\,square(2\pi\omega t)\ , a_2 = -2 - 20\,square(2\pi\omega t) \\ b_1 &= 5 + square(2\pi\omega t)\ , b_2 = 20 + 10\,square(2\pi\omega t) \end{aligned}$$

where *square* denotes the unity magnitude square wave function and ω is the plant switching frequency is Hz.

4.1 Control System Evaluation

In this section, an adaptive controller, which is based on the design procedure of Section 4. Let us choose the nominal gains $C = 100$ (feedback gain), adaptation filter gain $L = I$, where I is the identity matrix, then we have from Theorem 1 that the decay rate $\alpha = 1$ rad/sec. This should yield a settling time of at most 4 seconds for the closed loop system. Also the nominal value of the adaptation gain $\Gamma = 100I$ will be used. Whereas, a^* is chosen to be a constant vector a_{ave} taking the average values of the parameters a_1, a_2, b_1, b_2, i.e., when square functions are set to zero.

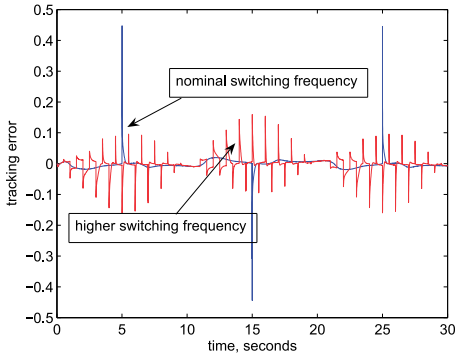


Fig. 1. Tracking error for different plant switching frequencies for developed adaptive controller.

Figure 1 shows the response of the modified adaptive controller for the output of the plant tracking a sinusoidal reference of amplitude 2 and frequency 0.3 rad/sec; the disturbance is set to zero for this case. The response follows the predicted theoretical behavior. The system responds to the corresponding impulse change in $\dot{a}$ and step change in a due to switching in plant parameter vector a with the error settling after exponentially decaying transient according to the system decay rate α. Whereas, by increasing the plant switching frequency, the same trend follows with no concern of instability. In fact, as the suggested by the bound in Theorem 2, plant parametric uncertainty and variation are inputs to the closed loop system. Therefore, increasing the frequency of this input, 6 rads/sec in this case, relative system bandwidth, 1 rads/ sec, will lead to attenuation of this input due to system roll-off as in linear systems. This explains why the tracking error is smaller for the higher switching frequency case.
Figure 2 shows the effect of different choices of the additional parameter estimate a^* for the nominal case of Figure 1. The figure shows that the average tracking error is larger when $a^* =$

$10\,a_{ave}$ and $a^* = 100\,a_{ave}$, since it corresponds to a larger size of the input $a^* - a$, as predicted by the bound of Theorem 2. The third case in Figure 2 shows the effect of switching the choice of a^* starting from a $a^* = 100\,a_{ave}$ to $a^* = 10\,a_{ave}$ at $t = 8$ seconds. Again, the response is that due to step changes in input $a^* - a$ with the transition between these two response takes place within the estimated settling time of 4 seconds based on a designed for decay rate of $\alpha = 1$ rads/ sec. This is a key capability that can be utilized in practice to perform robust and stable gain scheduling and online controller adjustments.

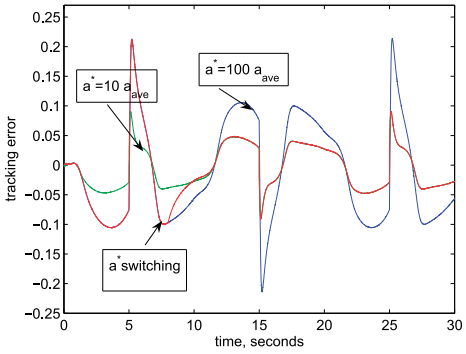


Fig. 2. Effect of parameter estimate a^* on tracking error for developed adaptive controller.

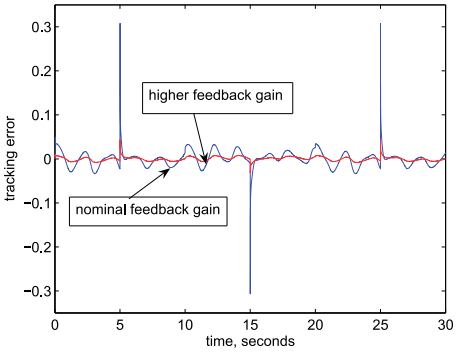


Fig. 3. Effect of feedback gain on tracking error for developed adaptive controller.

Next, Figures 3 and 4 will include the addition of a sinusoidal disturbance $d = 50\sin(\pi t)$ to the nominal case discussed above for switching frequency $\omega = 0.1$ Hz. Figure 3 displays the response of the nominal case of Figure 1 with the addition of a sinusoidal disturbance $d = 50\sin(\pi t)$, which introduces a clear sinusoidal content to the tracking error. Whereas, increasing feedback gain, which corresponds to matrix C in Theorem 1, significantly reduces the tracking error due to both plant switching (jumps and other steady errors) as well as the disturbance-induced error. This is consistent with the discussion in Section IV.B in that increasing system bandwidth α (via feedback gain) attenuates total input(disturbance an parametric uncertainties and variations) as well as speeds up the system bandwidth.

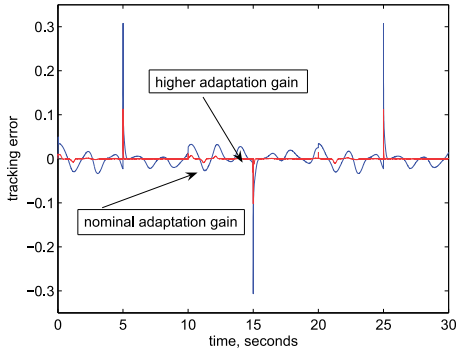


Fig. 4. Effect of adaptation gain Γ on tracking error for developed adaptive controller.

Whereas, Figure 4 considers the same situation in Figure 3 but with increasing adaptation gain instead of feedback gain. Again similar performance improvements are achieved along the lines of the bound in Theorem 1 yet without increasing system bandwidth.
Figures 2-4 show that error can be reduced by adjusting a^*, increasing feedback and adaptation gains, with different levels of effectiveness relative to disturbances, parametric uncertainty, and variation in accordance with the discussion in Section 3. The important message from this case study is not only that the developed control methodology can handel systems with large and rapid switching dynamics but also that this approach yields systematic and practical means to improve performance that follow the developed theory.

4.2 Comparison with Other Techniques

Finally, let us compare the system's response with the developed adaptive controller to other adaptive control techniques. We consider the same system of Section 5.1 with switching frequency $\omega = 1$ Hz case. The system is required to follow a constant reference of amplitude 2. First consider a non-adaptive backstepping controller, where the parameter estimate $\hat{a}$, in the developed control scheme of is replaced with a fixed value $\hat{a} = a_{ave}$. Figure 5 shows that the non-adaptive controller yields an unstable closed loop despite using the same assumed value of plant parameter vector, which has been used by the modified adaptive controller with $a^* = a_{ave}$.
Next, Figure 6 shows the response of the parameter estimates $\hat{a}$, when the equivalent standard adaptive controller, Equation (3), is used. This corresponds to setting $L = 0$ in the modified adaptive controller of Equation (6). In this case, some of the parameter estimates $\hat{a}$ grow unbounded, which could yield an unstable system in practical implementation. This is a known issue with standard adaptive control in the presence of parameter variations or even disturbances, which is usually referred to as parameter drift (6). In contrast, the modified adaptive controller for the same situation maintains bounded parameter estimates due to ISS stability of the closed loop, see Figure 7.
The poor robustness of standard adaptive controllers with respect to time varying parameters and disturbances has lead to modifying the adaptation law by robust adaptation laws such as deadzone, projection, and leakage modifications (6). Although there have not been any results reporting guaranteed stability and performance characteristics for rapidly varying switching

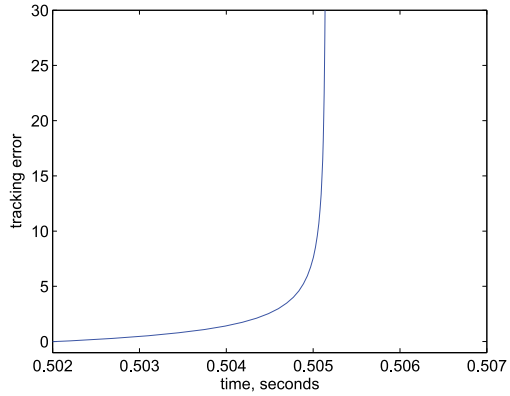


Fig. 5. Tracking error for non-adaptive backstepping controller with $\hat{a} = a_{ave}$.

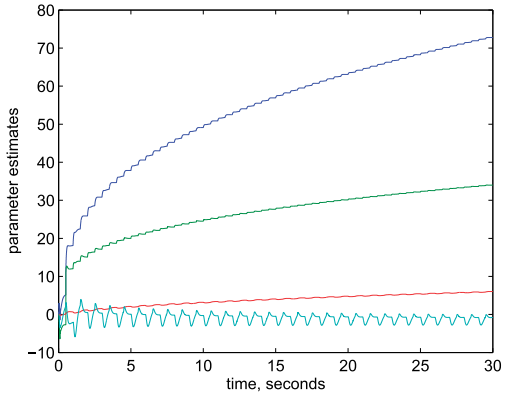


Fig. 6. Parameter estimates $\hat{a}$ for standard adaptive controller with $L = 0$.

systems using these techniques, we will compare the leakage-based modification developed in this chapter with deadzone and projection modifications.
A deadzone modification to the standard adaptation law of Equation (3) can be given by:

$$\dot{\hat{a}} = \begin{cases} f_a(x_m, \hat{a}, y_r, t) & \text{if } \|e\| > \epsilon \\ 0 & \text{otherwise} \end{cases}$$

This simply means to turn off the adaptation when the tracking error is less than some acceptable threshold ϵ. Figure 8 compares the modified adaptive controller with $a^* = a_{ave}$ to an equivalent deadzone adaptive controller with the same adaptation gain $\Gamma = 10000I$, where I is the identity matrix, and a deadzone threshold of $\epsilon = 0.3$. In this case, the modified adaptive controller outperforms the deadzone adaptive controller in the tracking error. Furthermore, when attempting to reduce the size of the tracking error threshold for the deadzone, ϵ, to allow for improvement in tracking error, the parameter estimates grew unboundedly as in the

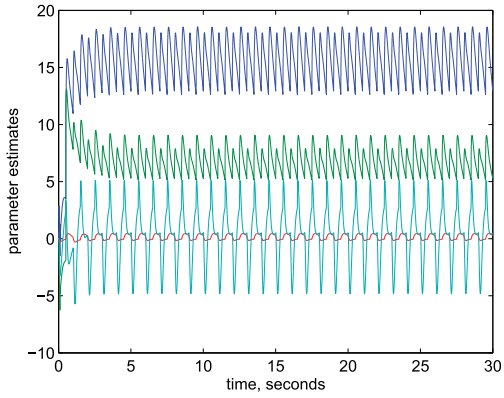


Fig. 7. Parameter estimates $\hat{a}$ for developed adaptive controller.

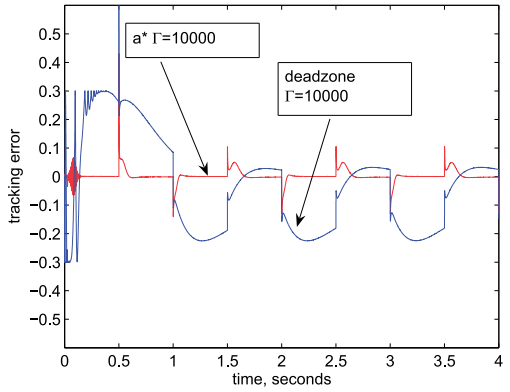


Fig. 8. Tracking error comparison for developed adaptive controller and a deadzone adaptive controller.

standard adaptive controller case of Figure 6. This is expected as the deadzone adaptive controller approaches that of a standard adaptive controller as $\epsilon \to 0$. Another limitation to the deadzone controller is the lack of systematic dependence on control parameters such as the adaptation gain Γ unlike the modified adaptive controller. Figure 9 shows how increasing the adaptation gain from $\Gamma = 100I$ to $\Gamma = 10000I$ does not necessarily improve tracking but rather yields reduction and increase in tracking at different times and of different signs. This is contrasted with the modified adaptive controller when tested under the same conditions, Figure 10, where a clear reduction in tracking error is observed with increasing Γ, in accordance with the scaling relationship in Section 3.

Next, we consider a parameter projection modification to the standard adaptive controller of

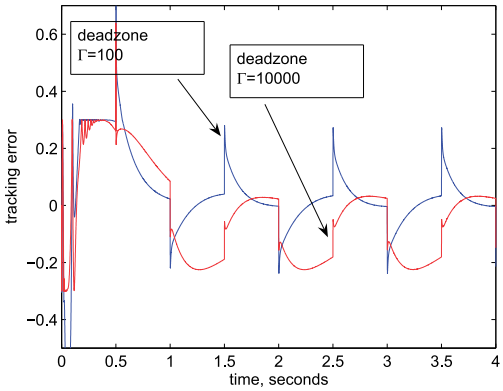


Fig. 9. Effect of adaptation gain Γ on tracking error for deadzone adaptive controller.

Equation (3). The projection modification (6) used here is given by:

$$\dot{\hat{a}} = \begin{cases} f_a & \text{if } \|\hat{a}\| \leq M \text{ or } \hat{a}^T f_a \leq 0 \\ f_a - \frac{\hat{a}\hat{a}^T}{\|\hat{a}\|^2}\left(\frac{\|\hat{a}\|^2 - M^2}{M^2}\right) f_a & \text{otherwise} \end{cases}$$

Which uses an assumed bound on parameters $\|a\| \leq M$. This assumption is critical to projection algorithms. Figure 11 shows the tracking error growing unbounded when a projection algorithm was implemented with a tight bound $M = 1$. In this case, the assumed bound on parameters was too tight as soon as the system switched to a different mode leading to instability. This is in contrast to the developed adaptive controller, which does not require such information to guarantee stability. This is the case as the assumed parameter vector a^* only affects the size of tracking error for a given choice of control gains.
Nevertheless, it was possible to obtain a choice for the projection bound, $M = 10$, where the system remained stable. Figure 12 compares the tracking error for this projection adaptive controller and the developed adaptive controller with $a^* = a_{ave}$ for the same adaptation gain. Again, the developed adaptive controller achieved smaller tracking error. As was the case with deadzone controller, the projection controller does not display the systematic dependence on the adaptation gain Γ unlike the proposed adaptive controller, see Figure 13. This is the case since both projection and deadzone modification do not achieve a clear bound due to ISS stability as that in Theorem 1. In fact, most results using such techniques to deal with disturbances or parameter variations only conclude boundedness. In this case, such a conclusion is of very little practical importance if the error can not be reduced to an acceptable level by increasing the adaptation gain or using a better nominal estimate of the plant parameters as with using a^* in the proposed adaptation law, see Figure 2.

5. Conclusions

A methodology for robust adaptive control design for a class switched nonlinear systems is presented. Under extensions of typical adaptive control assumptions, a leakage-type adap-

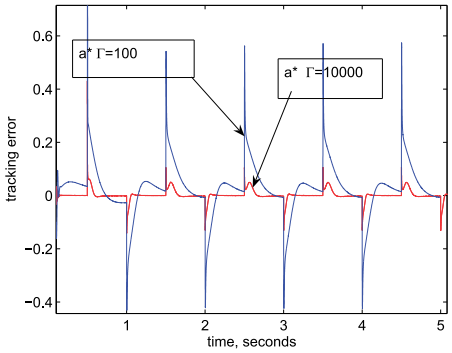


Fig. 10. Effect of adaptation gain Γ on tracking error for developed adaptive controller.

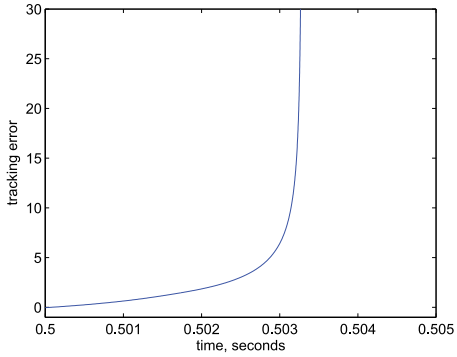


Fig. 11. Tracking error for projection adaptive controller with small parameter projection bound $M = 1$.

tive control scheme guarantees exponential and ISS stability with piecewise differentiable bounded plant parameters and piecewise continuous bounded disturbances without requiring a priori knowledge on such parameters. The effect of plant variation and switching is reduced to piecewise continuous and impulsive inputs acting on this ISS stable closed loop system. This yields a separation between robust stability and robust performance and clear guidelines for performance optimization via ISS bounds. The results are demonstrated through example simulations, which follow the developed theory and demonstrate superior robustness of stability and performance relative to non-adaptive and other adaptive methods such as projection and deadzone adaptive controllers. The authors believe that the use of these type of robust adaptive controllers is useful for switched systems even in the switched linear uncertainty free plant case, where stability with switched linear feedback is difficult to guarantee based on currently available tools.

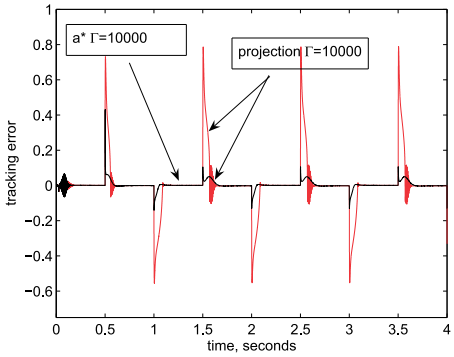


Fig. 12. Tracking error comparison for developed adaptive controller and a projection adaptive controller with large parameter projection bound $M = 10$.

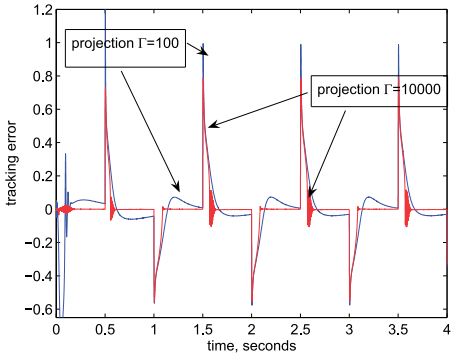


Fig. 13. Effect of adaptation gain Γ on tracking error for projection adaptive controller with large parameter projection bound $M = 10$.

6. References

[1] Anderson, B.D.O., Brinsmead, T.S., Liberzon, D., and Morse,A.S. Multiple model adaptive control with safe switching. *Int. J. Adaptive Control Signal Process.*, 15, 445-470, 2001.

[2] Decarlo, R. A. , Branicky, M. S., Pettersson, S., and B. Lennartson. Perspectives and Results on the Stability Stabilizability of Hybrid Systems, *Proceedings of The IEEE*, 88 (7), 1069-1082, July 2000.

[3] El Rifai, K., El Rifai, O., and K. Youcef-Toumi. On Robust Adaptive Switched Control, *American Control Conference 2005*, Portland, OR.

[4] El Rifai, and K. Youcef-Toumi. On Robust Adaptive Control of of a Class of Switched Systems, *American Control Conference 2008*, Seattle, WA.

[5] Ikhouane, F. and M. Krstic. Robustness of the Tuning Functions Adaptive Backstepping Design for Linear Systems.*IEEE transactions on Automatic Control*, 43 (3), 431-437, 1998.

[6] Ioannou P. and J. Sun. Robust Adaptive Control, *Prentice-Hall*, 1996.
[7] Kalkkuhl, J., Johansen, T.A., and J. Ldemann. Improved Transient Performance of Nonlinear Adaptive Backstepping Using Estimator Resetting Based on Multiple Models, *IEEE transactions on Automatic Control*, 47(1), 136-140, 2002.
[8] Khalil, H. Nonlinear Systems, 2nd Ed.,*Prentice-Hall*, 1995.
[9] Krstic, M., Kanellakopoulos, I., and P. V. Kokotovic. Nonlinear and Adaptive Control Design, *Wiley*, 1995.
[10] Liberzon, D., and A.S. Morse (1999). Basic Problems in Stability and Design of Switched Systems. *IEEE Control Systems Magazine*, 19, 59-70.
[11] Liberzon, D. Switching in Systems and Control, *Birkhauser*, 2003.
[12] Lu, B., and F. Wu. Control Design of Switched LPV systems Using Multiple Paramter-Dependent Lyapunov Functions, *American Control Conference*, Boston, MA, 2004.
[13] Marino, R. and P. Tomei. Robust Adaptive State-Feedback Tracking for Nonlinear Systems. *IEEE Transactions on Automatic Control*, 43(1), 84-89, 1998.
[14] Marino, R. and P. Tomei. An Adaptive Output Feedback Control for a Class of Nonlinear Systems with Time-Varying Parameters. *IEEE Transactions on Automatic Control*, 44 (11), 2190-2194, 1999.
[15] Marino, R. and P. Tomei. Control of linear time-varying systems. *Automatica*, 39, 651659, 2003.
[16] Narendra, K.S., Driollet, O.A., Feiler, M. , and K. George. Adaptive control using multiple models, switching and tuning.*Int. J. Adaptive Control Signal Process.*, 17,2003.
[17] Shorten, R.N., K.S. Narendra, and O. Mason (2003). A Result on Common Quadratic Lyapunov Functions. *IEEE Transactions on Automatic Control*, 48(1), 110-113.
[18] Shorten, R.N., and K.S. Narendra (2003). On Common Quadratic Lyapunov Functions for Pairs of Stable LTI Systems Whose System Matrices Are in Companion Form . *IEEE Transactions on Automatic Control*, 48(4), 618-621.
[19] Slotine, J.J.E., and Li, W. (1991). Applied Nonlinear Control, *Prentice-Hall*.
[20] Stilwell, D.J. and W.J. Rugh. Stability and L-2 Gain Properties of LPV Systrems.*Automatica*, 38 (9), 1601-1606, 2002.
[21] van der Schaft, A. and J. M. Schumacher. An Introduction to Hybrid Dynamical Systems, *Springer Verlag*, 2000.
[22] Zhai, G. Hu, B. Yasuda, K. and A. N. Michel. Disturbance Attenuation Properties of Time-Controlled Switched Systems, *Journal of the Franklin Institute*, 338 (7), 765-779, 2001.
[23] Zhang, D., L. Adaptive Control of Switched Systems. *IEEE Conference on Decision and Control*, Maui, Hawaii, USA, December 2003.
[24] Zhang, Y., Fidan, B., and P.Ioannou. Backstepping Control of Linear Time-Varying Systems With Known and Unknown Parameters. *IEEE Transactions on Automatic Control*, 48 (11), 1908-1925, 2003.
[25] Zhendong S., S.S. Geb. Analysis and synthesis of switched linear control systems. *Automatica*, 41, 181-195, 2005.
[26] Zhivoglyadov, P. V. , Middleton, R. H., and M. Fu, Localization Based Switching Adaptive Control for Time-Varying Discrete-Time Systems.*IEEE Transactions on Automatic Control*, 45 (4), 752-755, 2000.

3

Controller Synthesis for a Class of State-Dependent Switched Nonlinear Systems

Jenq-Lang Wu
Department of Electrical Engineering, National Taiwan Ocean University
Taiwan, R. O. C.

1. Introduction

In the last decade, switched systems have gained much attention since a large class of practical systems can be modeled as switched systems (see e.g., (Dayawansa & Martin, 1999)) and there exist systems that cannot be asymptotically stabilized by a single continuous feedback control law (see e.g., (Brockett, 1983)). For stability analysis of switched systems, many interesting results have been proposed in the literature, see e.g., (Narendra & Balakrishnan, 1994; Johansson & Rantzer, 1998; Branicky, 1998; Ye et al., 1998; Dayawansa & Martin, 1999; Liberzon et al., 1999; Skafidas et al., 1999; Mancilla-Aguilar, 2000; and Chatterjee & Liberzon, 2007). Another topic is the derivation of stabilizing switching rules for switched systems, see et al., (Peleties & DeCarlo, 1991; Liberzon & Morse, 1999; and Pettersson, 2003). For feedback controller synthesis of switched control systems (with input signals), most of the proposed results consider the linear subsystems case, see e.g., (Daafouz et al., 2002; Sun & Ge, 2003; Petterson, 2004; Hespanha & Morse, 2004; and Seatzu et al., 2006). Only a few results have been proposed for feedback controller synthesis of switched nonlinear control systems, see e.g., (Sun & Zhao, 2001; El-Farra et al., 2005; Wu, 2008; and Wu, 2009). In (El-Farra et al., 2005), an integrated synthesis of feedback controllers together with switching laws has been proposed. In (Sun & Zhao, 2001), a *common control Lyapunov function* (CCLF) approach has been introduced also for constructing feedback control laws together with switching signals for switched nonlinear control systems. The concept of CCLF is motivated by the *control Lyapunov function* approach (see, e.g., (Artstein, 1983) and (Sontag, 1983 and 1989)) for designing stabilizing feedback laws for (non-switched) nonlinear systems. In (Wu, 2008), for switched nonlinear control systems under arbitrarily switching, conditions for the existence of CCLFs has been derived and a globally uniformly asymptotically stabilizing feedback law has been proposed. However, no systematical approaches have been provided for constructing CCLFs. Moreover, the obtained feedback law is complicated. In (Wu, 2009), for switched nonlinear control systems, which arbitrarily switching between a set of subsystems in strict feedback form, the backstepping approach (see e.g., (Krstic et al., 1995) and (Sepulchre et al., 1997)) has been employed to construct CCLFs, and a simpler stabilizing feedback law has been proposed.

However, till now, few results have been reported in the literature about stabilizing feedback controllers design for state-dependent switched nonlinear control systems. The

purpose of this chapter is to give a constructive approach for this problem. The state space is partitioned, by a set of switching surfaces, into several operation regions. In each of these regions, a nonlinear dynamical system (in feedback linearizable form) is given. Whenever the state trajectory passes a switching surface, a new dynamical model dominates the system's behavior. That is, the switching signal is state-dependent and predetermined. For the stability analysis of state-dependent switched systems, a *common Lyapunov function* for all subsystems is easier to develop but too conservative, see e.g., (Liberzon, 2003). It is known that the *multiple Lyapunov function* approach is a less conservative method. Based on the concepts of *multiple Lyapunov functions* and *control Lyapunov functions*, this chapter introduces a *switched control Lyapunov function* (SCLF) approach for designing stabilizing feedback controllers for state-dependent switched nonlinear control systems. It should be emphasized that the derivation of CCLFs or SCLFs for switched nonlinear control systems is an open problem unless the systems are in some particular form. Therefore, in this chapter we restrict our attention to switched nonlinear systems in *feedback linearizable* form for the reason that, in this case, SCLFs can be chosen as piecewise quadratic form and can be obtained by solving bilinear matrix inequalities (BMIs) with equality constraints. Although the considered systems are in feedback linearizable form, we do not use the feedback linearization technique in the design procedure. We show that the considered stabilization problem for switched nonlinear control systems can be solved by directly solving a matrix problem. We will show that an explicit stabilizing switched feedback law, based on the Sontag's formula (see (Sontag, 1989)), can be easily derived once a SCLF has been obtained.

NOTATIONS: That $A \setminus B$ is the set of all elements which belong to set A but not belong to set B; $A \cup B$ is the union of sets A and B; $A \cap B$ is the intersection of sets A and B; clA is the closure of set A; $\overline{A}$ is the boundary of set A; $IntA$ is the interior of set A (i.e., $IntA = clA \setminus \overline{A}$); $P > 0$ ($P < 0$) means that the matrix P is positive (negative) definite; $P \geq 0$ ($P \leq 0$) means that the matrix P is positive (negative) semidefinite; ϕ denotes an empty set; $\forall$ means "for all".

2. Problem Formulation and Preliminaries

The intension of this section is to present some preliminaries and to explicitly formulate the problem to be solved.

2.1 Switched nonlinear control systems

In this chapter we are focused on switched nonlinear systems with input signals:

$$\dot{x} = A_{\sigma(x)}x + C_{\sigma(x)}f_{\sigma(x)}(x) + B_{\sigma(x)}g_{\sigma(x)}(x)u\ ,\ \sigma(x) \in \{1,...,q\} \tag{1}$$

where $x \in R^n$ is the system state, $u \in R^m$ is the control input, $A_i \in R^{n\times n}$, $C_i \in R^{n\times s}$, $B_i \in R^{n\times m}$, i=1,…,q, are constant matrices, $\sigma : R^n \mapsto \{1,...,q\}$ is a predetermined state-dependent switching signal, and $f_i(x) \in R^s$ and $g_i(x) \in R^{m\times m}$, i=1,…,q, are locally Lipchitz functions. Suppose that $f_i(0) = 0$, $g_i(x)$ is nonsingular for all $x \in \Omega_i$. Suppose also that

$$N(C_i^T) \subset N(B_i^T) \text{ for each } i \in \{1,...,q\}\ . \tag{2}$$

Define the index set $I_S = \{1,...,q\}$. Associated with the considered switched control system (1), a family of subsystems is defined:

$$\dot{x} = A_i x + C_i f_i(x) + B_i g_i(x)u,\ i \in I_S. \tag{3}$$

By (2), it can be seen that the subsystems in (3) are feedback linearizable (see Khalil, 1996).

2.2 State space partition

Specially in this chapter, the state space is partitioned into q regions Ω_i, i=1,…,q, given by quadratic forms:

$$\Omega_i \equiv \left\{ x \in R^n \middle| x^T Q_i x \geq 0 \right\},\ i=1,\ldots,q, \tag{4}$$

for some symmetric matrices $Q_i \in R^{n\times n}$, i=1,…,q. Let $Int\Omega_i \cap Int\Omega_j = \phi$ if $i \neq j$, and $\Omega_1 \cup \cdots \cup \Omega_q = R^n$. That is, the overlap between two adjacent regions is the boundary between these two regions. The i-th subsystem of (3) can be active only in part of the state space, specified by region Ω_i. Define the adjacent index set

$$I_A = \left\{ \{i,j\} \in I_S \times I_S \middle| \Omega_i \cap \Omega_j \setminus \{0\} \neq \phi \right\}. \tag{5}$$

That is, if $\{i,j\} \in I_A$, then Ω_i and Ω_j are adjacent regions and thus a switching region S_{ij} is defined:

$$S_{ij} \equiv \left\{ x \in R^n \middle| x^T (Q_i - Q_j) x = 0 \right\} \tag{6}$$

In fact, $S_{ij} = \Omega_i \cap \Omega_j$. Switches of the i-th subsystem into the j-th subsystem (or, switches of the j-th subsystem into the i-th subsystem) can occur only in the region S_{ij}. Note that in (Pettersson, 2004), for switched linear systems, the partition of state space is determined by the designer. That is, Q_i, i=1,…,q, are parameters to be determined. But in this chapter, we consider the case that they are predetermined.

2.3 Switching rule

In this chapter, we consider the case that the switching signal is state-dependent and is given by:

$$\sigma(x(t)) = i, \text{ if } x(t) \in Int\Omega_i, \text{ or } x(t) \in \overline{\Omega}_i \text{ and } \sigma(x(t^-)) = i. \tag{7}$$

By (7), the switching signal $\sigma(x)$ changes it value only if the state trajectory leaves one of the regions Ω_i, i=1,…,q. It holds constant value if the state trajectory keeps within a particular region (including its boundary).

2.4 Problem formulation

The goal of this chapter is to construct a state feedback law

$$u = h_{\sigma(x)}(x) \tag{8}$$

to globally asymptotically stabilize the switched control system (1). That is, we want to find a feedback law (8) such that the closed-loop system

$$\dot{x} = A_{\sigma(x)}x + C_{\sigma(x)}f_{\sigma(x)}(x) + B_{\sigma(x)}g_{\sigma(x)}(x)h_{\sigma(x)}(x) \tag{9}$$

becomes globally asymptotically stable under the switching rule (7).

2.5 Multiple Lyapunov functions

As stated in (Liberzon, 2003), *common Lyapunov function* approach will be too conservative for stability analysis of state-dependent switched systems. The *multiple Lyapunov function* approach will be less conservative. Here we briefly review the concept of *multiple Lyapunov functions*.

To analyze the stability of state-dependent switched system (9), for each $i \in I_S$, we first find a Lyapunov-like function $V_i(x)$, which vanishes at the origin and is positive for all $x \in \Omega_i \setminus \{0\}$, for the *i*-th subsystem of (9). System (9) is stable if, for each $i \in I_S$, the values of $V_i(x)$ at every switching instants, when we enter (switch into) the *i*-th subsystem, form a monotonically decreasing sequence.

However, using the *multiple Lyapunov function* approach in practically analyzing stability is difficult since, for verifying the monotonically decreasing property, one must have some information about the solutions of the switched systems (Liberzon, 2003). That is, one needs to know the values of suitable Lyapunov-like functions at switching times, which in general requires the knowledge of the state at these times. This is to be contrasted with the classical Lyapunov stability results, which do not require the knowledge of solutions, see (Liberzon, 2003). To simply the analysis procedure, an additional assumption that the *multiple Lyapunov function* is continuous on the boundaries between regions (i.e., switching surfaces) can be introduced. This assumption is conservative but leads to a simpler condition for verifying stability.

2.6 Switched control Lyapunov functions

Multiple Lyapunov function approach can be used to determine the stability of switched systems without input signals. However, it cannot tell us how to find a stabilizing feedback law (8) for the switched control system (1). Here we introduce the *switched control Lyapunov function* (SCLF) for feedback controller synthesis.

Definition 1: A switched function $V(x) \equiv V_{\sigma(x)}(x)$, which is differentiable in $Int\Omega_i$, for all $i \in I_S$, and continuous on S_{ij}, for all $\{i, j\} \in I_A$, is a SCLF of (1) if,

$$V(0) = 0 \tag{10}$$

$$V(x) > 0, \ \forall x \in R^n \setminus \{0\} \tag{11}$$

$$V(x) \to \infty, \text{ as } \|x\| \to \infty \tag{12}$$

and, along each possible nonzero solution of (1), a control signal u exists such that $V(x)$ monotonically decreases. ■

Similar to the statement about *multiple Lyapunov function* in the previous subsection, for simplifying the design procedure, we make the assumption that the SCLF is continuous everywhere. By Definition 1, if we can find a SCLF for the switched system (1), then for each solution of (1) we can derive a control signal u such that SCLF monotonically decreases. The next problems are how to derive SCLFs for (1) and how to develop stabilizing feedback controllers by the obtained SCLF.

3. Stabilizing Controller Synthesis

In this section we propose the main results, a sufficient condition for the existence of SCLFs for (1) and a stabilizing feedback law derived by the obtained SCLF.
We first define the *regional control Lyapunov functions* (*RCLFs*) for the subsystems in (3), which will be used to construct SCLFs for (1).
Definition 2: A differentiable function $V_i(x)$ is a Ω_i -RCLF for the i-th subsystem of (3) if,

$$V_i(0) = 0 \tag{13}$$

$$V_i(x) > 0\,,\ \forall x \in \Omega_i \setminus \{0\} \tag{14}$$

$$V_i(x) \to \infty\,, \text{as } x \in \Omega_i \text{ and } \|x\| \to \infty \tag{15}$$

and

$$\inf_{u \in R^m} \frac{\partial V_i(x)}{\partial x}\left(A_i x + C_i f_i(x) + B_i g_i(x) u\right) < 0\,,\ \forall x \in \Omega_i \setminus \{0\} \tag{16}$$

■

Define

$$a_i(x) \equiv \frac{\partial V_i(x)}{\partial x}\left(A_i x + C_i f_i(x)\right) \tag{17}$$

and

$$b_i(x) \equiv \frac{\partial V_i(x)}{\partial x} B_i g_i(x)\,. \tag{18}$$

By Definition 2, a differentiable function $V_i(x)$, satisfying (13)-(15), is a Ω_i -RCLF for the i-th subsystem if

$$\forall x \in \Omega_i \setminus \{0\}\,,\ b_i(x) = 0 \Rightarrow a_i(x) < 0\,. \tag{19}$$

Notice that if we can find a Ω_i -RCLF, $V_i(x)$, for the i-th subsystem, then for all $x \in \Omega_i \setminus \{0\}$ we can derive an u such that $a_i(x) + b_i(x)u < 0$.
Now we recall the S-procedure and the Finsler's lemma which will be used latter for deriving conditions for the existence of RCLFs.

Lemma 1 (*S-procedure*) (Boyd et al., 1994)*:* Let $P \in R^{n\times n}$ and $Q \in R^{n\times n}$ be symmetry. Then

$$x^T Px > 0 \text{ for all } x \neq 0 \text{ satisfies } x^T Qx \geq 0$$

if there exists a scalar $\rho \geq 0$ such that

$$P - \rho Q > 0 . \qquad \blacksquare$$

Lemma 2 (*Finsler's Lemma*) (Boyd et al., 1994): Consider a symmetric matrix $P \in R^{n\times n}$ and a matrix $N \in R^{n\times m}$, with *rank*(*N*)<*n*. The following statements are equivalent:

1) $x^T Px < 0 \quad \forall x \neq 0$ such that $N^T x = 0$
2) $\exists \mu \in R$ such that $P - \mu NN^T < 0$
3) $\exists L \in R^{m\times n}$ such that $P + L^T N^T + NL < 0$. $\blacksquare$

By the particular structure of the switched control system (1), RCLFs of the subsystems in (3) can be chosen as quadratic form and then can be obtained by solving bilinear matrix inequalities (BMIs). From Definition 2, a quadratic function $V_i(x) \equiv x^T P_i x$, with $P_i = P_i^T \in R^{n\times n}$, is a Ω_i-RCLF for the *i*-th subsystem of (3) if

$$x^T P_i x > 0 \text{ for all } x \neq 0 \text{ such that } x^T Q_i x \geq 0 \tag{20}$$

and

$$\inf_{u\in R^m} x^T P_i \left(A_i x + C_i f_i(x)\right) + x^T P_i B_i g_i(x) u < 0 \text{, for all } x \neq 0 \text{ such that } x^T Q_i x \geq 0 \tag{21}$$

We have the following result.

Theorem 1: There exists a quadratic Ω_i-RCLF for the *i*-th subsystem of (3) if there exist scalars $\rho_i \geq 0$ and $\eta_i \geq 0$, and matrices $L_i \in R^{m\times n}$ and $P_i = P_i^T \in R^{n\times n}$ satisfy the following bilinear matrix inequalities:

$$P_i - \eta_i Q_i > 0 \tag{22}$$

$$A_i^T P_i + P_i A_i + \rho_i Q_i + L_i^T B_i^T P_i + P_i B_i L_i < 0 \tag{23}$$

In this case, the quadratic function $V_i(x) \equiv x^T P_i x$ is a Ω_i-RCLF for the *i*-th subsystem of (3).

Proof: From (20), (22) and Lemma 1, it is clear that $V_i(x) > 0$ for all $x \in \Omega_i \setminus \{0\}$. Notice that

$$b_i(x) = \frac{\partial V_i(x)}{\partial x} B_i g_i(x) = 2x^T P_i B_i g_i(x) \tag{24}$$

and

$$a_i(x) = \frac{\partial V_i(x)}{\partial x}\left(A_i x + C_i f_i(x)\right) = 2x^T P_i A_i x + 2x^T P_i C_i f_i(x) \tag{25}$$

For $x \in \Omega_i \setminus \{0\}$ such that $b_i(x) = 0$, we have $P_i x \in N(B_i^T)$ since $g_i(x)$ is nonsingular. Then, $P_i x \in N(C_i^T)$ by (2) and therefore $x^T P_i C_i f_i(x) = 0$. From (23), for $x \in \Omega_i \setminus \{0\}$ satisfying $b_i(x) = 0$, we have

$$\begin{aligned} a_i(x) &= 2x^T P_i A_i x + 2x^T P_i C_i f_i(x) \\ &= 2x^T P_i A_i x \\ &< -x^T(\rho_i Q_i + L_i^T B_i^T P_i + P_i B_i L_i)x \\ &= -\rho_i x^T Q_i x \\ &\le 0 \end{aligned}$$

This proves the result by noting (19). ■

It should be noted that, by the conditions in Theorem 1, $V_i(x) = x^T P_i x$ is not a classical *control Lyapunov function* for the *i*-th subsystem. This is obvious since the solution P_i may not be positive definite.

By solving (22) and (23) for all $i \in I_S$, if all the subsystems in (3) have their RCLFs, $V_i(x) = x^T P x$, *i*=1,…,*q*, one might think that the switched function $V_{\sigma(x)}(x)$ is a SCLF for switched control system (1). However, this is not true since these RCLFs in general have different values on the switching surfaces. That is, $V_{\sigma(x)}(x)$ will be discontinuous on the switching surfaces. If we use $V_{\sigma(x)}(x)$ as a *control Lyapunov function* for (1), we can find control signal *u* such that $V_{\sigma(x)}(x)$ decreases between sequel switching times. However, $V_{\sigma(x)}(x)$ may increase at the switching instants (i.e., as the trajectories of (1) pass through the switching surfaces). In this case, the design of stabilizing feedback laws is difficult. To simply the design procedure, an additional continuity requirement is included for SCLFs. In the next theorem we introduce additional constraints in solving the matrix inequalities to guarantee the continuity of SCLFs on the switching surfaces. Moreover, a stabilizing feedback controller is given provided that a SCLF is obtained.

Theorem 2: There exists a piecewise quadratic SCLF for the switched control system (1) if there exist scalars $\rho_i \ge 0$ and $\eta_i \ge 0$, and matrices $L_i \in R^{m\times n}$ and $P_i = P_i^T \in R^{n\times n}$, *i*=1,2,…,*q*, and scalars δ_{ij}, for all $\{i, j\} \in I_A$, satisfy the following matrix inequalities and equalities:

$$P_i - \eta_i Q_i > 0,\ i=1,2,\ldots,q, \tag{26}$$

$$A_i^T P_i + P_i A_i + \rho_i Q_i + L_i^T B_i^T P_i + P_i B_i L_i < 0,\ i=1,2,\ldots,q, \tag{27}$$

$$P_i - P_j = \delta_{ij}(Q_i - Q_j),\ \text{for all } \{i, j\} \in I_A. \tag{28}$$

In this case, the function $V_{\sigma(x)}(x) = x^T P_{\sigma(x)} x$ is a SCLF for (1). In addition,

$$u(x) = h_{\sigma(x)}(x) \tag{29}$$

with ($k > 0$)

$$h_i(x) \equiv \begin{cases} -b_i^T(x)\dfrac{a_i(x) + k\sqrt{a_i^2(x) + \left(b_i(x)b_i^T(x)\right)^2}}{b_i(x)b_i^T(x)}, & \text{if } b_i(x) \neq 0 \\ 0, & \text{if } b_i(x) = 0 \end{cases} \tag{30}$$

is an asymptotically stabilizing feedback law for (1) under the switching rule (7).

Proof: By (28) it is clear that $x^T P_i x = x^T P_j x$ on S_{ij}, for all $\{i, j\} \in I_A$. That is, $V_{\sigma(x)}(x)$ is continuous in all state space. By Definition 1 and Theorem 1 and noting (26) and (27), it is obvious that $V_{\sigma(x)}(x) \equiv x^T P_{\sigma(x)} x$ is a SCLF for (1).

To show that (29) is a stabilizing feedback law, notice that, for each $i \in I_S$, if $x \in \Omega_i \setminus \{0\}$ (and $\sigma(x) = i$) such that $b_i(x) = 0$, we have

$$\begin{aligned} \dot{V}_{\sigma(x)}(x) = \dot{V}_i(x) &= a_i(x) + b_i(x)u \\ &= a_i(x) + b_i(x)h_i(x) \\ &= a_i(x) \\ &< 0. \end{aligned}$$

Moreover, if $x \in \Omega_i \setminus \{0\}$ (and $\sigma(x) = i$) such that $b_i(x) \neq 0$, then

$$\begin{aligned} \dot{V}_{\sigma(x)}(x) = \dot{V}_i(x) &= a_i(x) + b_i(x)u \\ &= a_i(x) + b_i(x)h_i(x) \\ &= -k\sqrt{a_i^2(x) + \left(b_i(x)b_i^T(x)\right)^2} \\ &< 0. \end{aligned}$$

That is, if no sliding motions occur on the switching surfaces, the closed-loop system is asymptotically stable since $\dot{V}_{\sigma(x)}(x) < 0 \ \ \forall x \neq 0$ (notice that the index set I_S is countable).

In the case that sliding motion occurs, we need to prove the stability of sliding motion. If a sliding motion occurs on S_{ij} for some $\{i, j\} \in I_A$, first suppose that $V_i(x) \leq V_j(x)$ for $x \in \Omega_i$ and $V_i(x) \geq V_j(x)$ for $x \in \Omega_j$ in a neighbourhood of S_{ij}. The existence of a sliding mode on S_{ij} is characterized by the inequalities

$$x^T(P_i - P_j)\left(A_i x + C_i f_i(x) + B_i g_i(x)h_i(x)\right) \geq 0 \tag{31}$$

$$x^T(P_i - P_j)\left(A_j x + C_j f_j(x) + B_j g_j(x)h_j(x)\right) \leq 0\,. \tag{32}$$

Since a sliding motion occurs on S_{ij}, σ is not uniquely defined on S_{ij}. So let $\sigma(x)=i$ without loss of generality (Liberzon, 2003). Along the corresponding Filippov solution, by (32), we have (for $\alpha \in (0,1)$)

$$\begin{aligned}\dot{V}_i(x) &= 2x^T P_i\left(\alpha\left(A_i x + C_i f_i(x) + B_i g_i(x) h_i(x)\right) + (1-\alpha)\left(A_j x + C_j f_j(x) + B_j g_j(x) h_j(x)\right)\right)\\ &= 2\Big(\alpha\left(x^T P_i A_i x + x^T P_i C_i f_i(x) + x^T P_i B_i g_i(x) h_i(x)\right)\\ &\quad + (1-\alpha)\left(x^T P_i A_j x + x^T P_i C_j f_j(x) + x^T P_i B_j g_j(x) h_j(x)\right)\Big)\\ &\le 2\Big(\alpha\left(x^T P_i A_i x + x^T P_i C_i f_i(x) + x^T P_i B_i g_i(x) h_i(x)\right)\\ &\quad + (1-\alpha)\left(x^T P_j A_j x + x^T P_j C_j f_j(x) + x^T P_j B_j g_j(x) h_j(x)\right)\Big)\\ &< 0, \text{ if } x \neq 0.\end{aligned}$$

That is, $V_i(x)$ monotonically decreases along the corresponding Filippov solution. Similarly, if $V_i(x) \ge V_j(x)$ for $x \in \Omega_i$ and $V_i(x) \le V_j(x)$ for $x \in \Omega_j$ in a neighbourhood of S_{ij}, we can show that $\dot{V}_j(x) < 0$ along the corresponding Filippov solution. This implies that the switched closed-loop system is still asymptotically stable even if sliding motions occur on the switching surfaces. ■

The formula in (30) is from the well known Sontag's formula (Sontag, 1989). We can also construct the feedback law by the Freeman's formula, see (Krstic et al., 1995).

Remark 1: In the case that the partition of state space is not predetermined and the switching signal is also a design parameter, by the covering property (Pettersson, 2004), some additional matrix inequalities must be included together with (26)-(28) to guarantee the existence of feedback controllers and switching laws for stabilizing the switched control systems. ■

4. An Illustrative Example

Consider the switched nonlinear control system

$$\dot{x} = A_{\sigma(x)} x + C_{\sigma(x)} f_{\sigma(x)}(x) + B_{\sigma(x)} g_{\sigma(x)}(x) u,\ \sigma(x) \in \{1,2\} \tag{33}$$

with the following system parameters:

$$A_1 = \begin{bmatrix} 1 & -5 \\ -1 & 0 \end{bmatrix},\ C_1 = \begin{bmatrix} 0 \\ -3 \end{bmatrix},\ B_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},\ f_1(x) = x_1^2 x_2,\ g_1(x) = -1 - x_2^2,$$
$$A_2 = \begin{bmatrix} 3 & -1 \\ 3 & 2 \end{bmatrix},\ C_2 = \begin{bmatrix} 2 \\ -2 \end{bmatrix},\ B_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix},\ f_2(x) = x_1 x_2 - |x_2|,\ g_2(x) = 1 + |x_1 - x_2|.$$

The state space is partitioned as $R^2 = \Omega_1 \cup \Omega_2$, where

$$\Omega_1 \equiv \left\{ x \in R^2 \middle| x^T Q_1 x \geq 0 \right\} \text{ and } \Omega_2 \equiv \left\{ x \in R^2 \middle| x^T Q_2 x \geq 0 \right\}$$

with

$$Q_1 = -Q_2 = \begin{bmatrix} -1 & 1.05 \\ 1.05 & 1 \end{bmatrix}.$$

That is, the switching region (boundaries between Ω_1 and Ω_2) is:

$$S_{12} \equiv \left\{ x \in R^2 \middle| x^T (Q_1 - Q_2) x = 0 \right\}$$

The switching signal is state-dependent and is given by:

$$\sigma(x(t)) = i \text{ , if } x(t) \in Int\Omega_i \text{ , or } x(t) \in \overline{\Omega}_i \text{ and } \sigma(x(t^-)) = i \text{ .} \tag{34}$$

From Theorem 2, we can verify that,

$$\eta_1 = \eta_2 = 0.1 \text{ , } \rho_1 = \rho_2 = 3.4 \text{ , } L_1 = \begin{bmatrix} 1 & 1 \end{bmatrix} \text{, } L_2 = \begin{bmatrix} -2 & 1 \end{bmatrix} \text{,}$$
$$\delta_{12} = 1 \text{ , } P_1 = \begin{bmatrix} 1.2 & 1.1 \\ 1.1 & 3.2 \end{bmatrix} \text{, and } P_2 = \begin{bmatrix} 3.2 & -1 \\ -1 & 1.2 \end{bmatrix}$$

satisfy (26)-(28). Therefore, there exist stabilizing feedback laws for the considered switched control system (33). By (24) and (25), we have

$$a_1(x) = 0.2x_1^2 - 16.2x_1x_2 - 11x_2^2 - (6.6x_1 + 19.2x_2)x_1^2x_2 \text{ ,}$$
$$a_2(x) = 13.2x_1^2 - 9.2x_1x_2 + 6.8x_2^2 + (16.8x_1 - 8.8x_2)(x_1x_2 - |x_2|) \text{ ,}$$
$$b_1(x) = -(2.2x_1 + 6.4x_2)(1 + x_2^2) \text{ ,}$$
$$b_2(x) = (8.4x_1 - 4.4x_2)(1 + |x_1 - x_2|) \text{ .}$$

Moreover, by (30), we have

$$h_1(x) \equiv \begin{cases} -b_1^T(x) \dfrac{a_1(x) + k\sqrt{a_1^2(x) + \left(b_1(x)b_1^T(x)\right)^2}}{b_1(x)b_1^T(x)}, & \text{if } b_1(x) \neq 0 \\ 0, & \text{if } b_1(x) = 0 \end{cases}$$

$$h_2(x) \equiv \begin{cases} -b_2^T(x) \dfrac{a_2(x) + k\sqrt{a_2^2(x) + \left(b_2(x)b_2^T(x)\right)^2}}{b_2(x)b_2^T(x)}, & \text{if } b_2(x) \neq 0 \\ 0, & \text{if } b_2(x) = 0. \end{cases}$$

Then

$$u(x) = h_{\sigma(x)}(x) \tag{35}$$

is a stabilizing feedback controller for (33) under the switching rule (34).

Fig. 1 shows the state trajectories of the closed-loop switched system starting from several different initial conditions with $k = 0.01$. Notice that sliding motions occur.

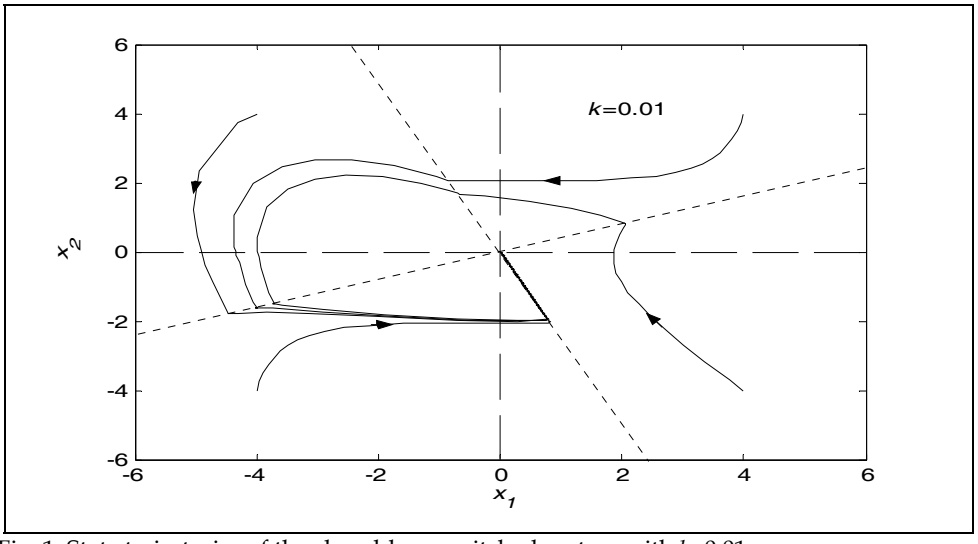


Fig. 1. State trajectories of the closed-loop switched system with k=0.01.

5. Conclusion

In this chapter, based on the use of switched control Lyapunov function approach, it has been shown that the design of stabilizing feedback laws for state-dependent nonlinear control systems in feedback linearizable form can be achieved by solving matrix problems. An example is given to illustrate the success of the method. However, solving the resultant bilinear matrix inequalities with equality constraints is not easy.

Further research topics include the development of feasible and efficient algorithms for solving the resultant matrix problem, the extension of the proposed approach to nonlinear control systems in some more general forms, and the search of stabilizing feedback laws to guarantee the non-existence of sliding motions.

6. Acknowledgements

The author would like to thank the National Science Council of the Republic of China for financially supporting this research under Contract No. NSC 97-2221-E-019-026-MY2.

7. References

Artstein, Z. (1983). Stabilization with relaxed control. *Nonlinear Analysis, Theory, Methods & Applications*, Vol. 7, pp. 1163-1173.

Boyd, S.; Ghaoui, L. E.; Feron, E. & Balakrishnan. V. (1994). *Linear Matrix Inequalities in Systems and Control Theory*, volume 15 of SIAM Studies in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia.

Branicky, M. S. (1998). Multiple Lyapunov functions and other analysis tools for switched and hybrid systems. *IEEE Trans. Automatic Control*, Vol. 43, pp. 475-482.

Brockett, R. W. (1983). Asymptotic stability and feedback stabilization. in *Differential Geometric Control Theory*, Birkhauser, Boston, pp. 181-191.

Chatterjee D. & Liberzon, D. (2007). On stability of randomly switched nonlinear systems. *IEEE Trans. Automatic Control*, Vol. 52, No. 12, pp. 2390-2394.

Daafouz, J.; Riedinger, P. & Iung, C. (2002). Stability analysis and control synthesis for switched systems: A switched Lyapunov function approach. *IEEE Trans. Automatic Control*, Vol. 47, pp. 1883-1887.

Dayawansa, W. P. & Martin, C. F. (1999). A converse Lyapunov theorem for a class of dynamical systems which undergo switching. *IEEE Trans. Automatic Control*, Vol. 44, pp. 751-760.

El-Farra, N. H.; Mhaskar, P. & Christofides, P. D. (2005). Output feedback control of switched nonlinear systems using multiple Lyapunov functions. *Proceedings of the American Control Conference*, Portland, pp. 3792-3799.

Hespanha, J. & Morse, A. S. (2004). Switching between stabilizing controllers. *Automatica*, Vol. 38, pp. 1905–1917.

Johansson, M. & Rantzer, A. (1998). Computation of piecewise quadratic Lyapunov functions for hybrid systems. *IEEE Control System Magazine*, Vol. 43, pp. 555-559.

Khalil, H. (1996). *Nonlinear Systems*, 2nd edition. Prentice_Hall.

Krstic, M.; Kanellakopoulos, I. & Kokotovic, P. (1995), *Nonlinear and Adaptive Control Design*, Wiley-Interscience.

Liberzon, D. (2003). *Switching in Systems and Control*. Birkhauser.

Liberzon, D.; Hespanha, J. P. & Morse, A. S. (1999). Stability of switched systems: a Lie-algebraic condition. *System & Control Letters*, Vol. 37, pp. 117-122.

Liberzon, D. & Morse, A. S. (1999). Basic problems in stability and design of switched systems. *IEEE Control System Magazine*, Vol. 19, No. 5, pp. 59-70.

Mancilla-Aguilar, J. L. (2000). A condition for the stability of switched nonlinear systems. *IEEE Trans. Automatic Control*, Vol. 45, No. 11, pp. 2077-2079.

Narendra, K. S. & Balakrishnan, J. (1994). A common Lyapunov function for stable LTI systems with com-muting A-matrices. *IEEE Trans. Automatic Control*, Vol. 39, pp. 2469-2471.

Peleties, P. & DeCarlo, R. (1991). Asymptotic stability of m-switched systems using Lyapunov-like functions. *Proceedings of the 1991 American Control Conference*, Boston, Massachusetts USA, pp. 1679-1684.

Petterson, S. (2003). Synthesis of Switched Linear Systems. *Proceedings of the 42nd IEEE Conference on Decision & Control*, Maui, Hawaii USA, pp. 5283-5288.

Petterson, S. (2004). Controller Design of Switched Linear Systems. *Proceedings of the 2004 American Control Conference*, Boston, Massachusetts USA, pp. 3869-3874.

Seatzu, C.; Corona, D.; Giua, A. & Bemporad, A. (2006). Optimal control of continuous-time switched affine systems. *IEEE Trans. Automatic Control*, Vol. 51, No. 5, pp. 726-741.

Sepulchre, R.; Jankovic, M. & Kokotovic, P. (1997), *Constructive Nonlinear Control*, Springer.

Skafidas, E.; Evans, R. J.; Savkin, A. V. & Peterson, I. R. (1999). Stability results for switched controller systems. *Automatica*, Vol. 35, pp. 553-564.

Sontag, E. D. (1983). A Lyapunov-like characterization of asymptotic controllability. *SIAM J. Control and Optimization*, Vol. 21, pp. 462-471.

Sontag, E. D. (1989). A 'universal' constructive of Artstein's theorem on nonlinear stabilization. *Systems & Control Letters*, Vol. 12, pp. 542-550.

Sun, H. & Zhao, J. (2001). Control Lyapunov functions for switched control systems. *Proceedings of the 2001 American Control Conference*, Arlington, June 25-27, pp. 1890-1891.

Sun, Z. & Ge, S. S. (2003). Dynamic output feedback stabilization of a class of switched linear systems. *IEEE Trans. Circuits Systems: I*, Vol. 50, pp. 1111-1115.

Wu, J. L. (2008). Feedback stabilization for multi-input switched nonlinear systems: two subsystems case. *IEEE Trans. Automatic Control*, Vol. 53, No. 4, pp. 1037-1042.

Wu, J. L. (2009). Stabilizing controllers design for switched nonlinear systems in strict-feedback form. *Automatica*, Vol. 45, No. 4, pp. 1092-1096.

Ye, H.; Michel, A. N. & Hou, L. (1998). Stability theory for hybrid dynamical systems. *IEEE Trans. Automatic Control*, Vol. 43, pp. 461-474.

4

Robust H_∞ Control for Linear Switched Systems with Time Delay

Yan Li, Zhihuai Li and Xinmin Wang
Northwestern Polytechnical University
P. R. China

1. Introduction

Switched control has been applied widely in the intelligent robots, aerospace and aeronautics engineering and wireless communications. In this chapter, the robust $H\infty$ control for linear uncertain switched systems with time delay is studied.

Switched systems with time delay include the system with single time delay and the system with multiple time delays. Linear switched systems with time delay can be described as follows:

$$\begin{aligned} \dot{x}(t) &= A_{\sigma(t)}x(t) + \sum_{j=1}^{N} A_{dj\sigma(t)}x(t-\tau_j) + B_{\sigma(t)}u(t) + B_{1\sigma(t)}w(t) \\ z(t) &= C_{\sigma(t)} + x(t) + \sum_{j=1}^{N} C_{dj\sigma(t)}x(t-\tau_j) + D_{\sigma(t)}u(t) + B_{2\sigma(t)}w(t) \\ x(t) &= \phi(t), t \in (-\max(\tau_j), 0) \end{aligned} \tag{1}$$

where $x(t) \in R^n$ is the system state vector, $u(t) \in R^{m1}$ is the input vector, $z(t) \in R^{m2}$ is the output vector, $w(t) \in l_2$ is the disturbance vector, $\sigma(t) : [0, \infty) \to M = \{1, 2, 3..., m\}$ is the switching signal, $A_{\sigma(t)}$, $A_{dj\sigma(t)}$, $B_{\sigma(t)}$, $B_{1\sigma(t)}$, $C_{\sigma(t)}$, $C_{dj\sigma(t)}$, $D_{\sigma(t)}$, $B_{2\sigma(t)}$ are known constant matrices, $\phi(t)$ represents the initial condition of the system, τ_j represents the time delay. For the system (1), if N=1, it is a switched system with single time delay, otherwise it is a switched system with multiple time delays.

The state feedback control for switched systems can be designed with memory or without memory.

For the switched system (1), the state feedback control can be designed as follows:

$u(t) = K_i x(t)$ is the state feedback control without memory;

$u(t) = K_i x(t) + \sum_{j=1}^{N} K_j x(t-\tau_{ij})$ is the state feedback control with memory.

Compared with the results on the stability of switched systems, research on the $H\infty$ control for switched systems is not adequate yet. Attentions have been attracted to the $H\infty$ control

for switched systems since 1998, when Hespanha considered the problem firstly. Similar to the stability problem, the $H\infty$ control problem can be classified into:

Problem A. The $H\infty$ control under arbitrary switching signal;

Problem B. The $H\infty$ control under a certain switching signal.

Problem A means the internal stability and the L_2 gain of the switched systems are independent of the switching signal. Problem A is usually solved through the common Lyapunov method which is conservative in that the common Lyapunov function is not easy to choose.

Wu and Meng (Wu & Meng, 2009) studied $H\infty$ model reduction for continuous-time linear switched systems with time-varying delay. By applying the average dwell time approach and the piecewise Lyapunov function technique, delay delay-dependent and delay-independent sufficient conditions are proposed in terms of linear matrix inequality (LMI) to guarantee the exponential stability and the weighted $H\infty$ performance for the error system. Zhang and Liu (Zhang & Liu 2008) studied the problem of delay-dependent robust $H\infty$ control for switched systems with disturbance and time-varying structured uncertainties. A sufficient condition ensuring the robust stabilization and $H\infty$ performance under arbitrary switching laws was obtained based on the Lyapunov function and Finslerpsilas lemma. Xie et al. (Xie *et al., 2004)* proposed conditions for uniformly quadratic stability for uncertain switched systems based on common Lyapunov method and LMI formulation. Fu et al. (Fu et al., 2007) proposed a the sufficient condition for the design of dynamic output feedback control of switched systems based on the common Lyapunov function approach and convex combination technique. Song et al. (Song et al., 2007) present the switching law and robust $H\infty$ control design for a class of discrete switched systems with time-varying delay. Song (Song et al., 2006) also studied a class of uncertain discrete switched systems with time delay. The switching law and the $H\infty$ controller are given based on the Multi-Lyapunov Function method. Ma et al. (Ma et al., 2006) proposed an $H\infty$ controller with memory for discrete switched systems with time delay.

In this chapter, the robust $H\infty$ control based on multi-Lyapunov-Function approach and LMI formulation for general linear switched systems with time delay is first introduced. The results are then extended to robust $H\infty$ control without and with memory for uncertain linear switched systems with time-varying delay. Suppose all sub-systems are not robust stable, a sufficient condition for system stabilization with $H\infty$ bound is given, as well as the design algorithm for the robust $H\infty$ switched control and the switching law. The simulation results show the effectiveness of the methods.

2. Robust H_∞ stability and stabilization for linear switched systems

Consider the following linear switched system:

$$\begin{aligned} \dot{x} &= A_i x + B_i w \\ z &= C_i x \end{aligned} \tag{2}$$

where , $x(t) \in R^n$ is the state vector, $z(t) \in R^{m2}$ is the output vector, $w(t) \in l_2$ is the disturbance vector, A_i, B_i, C_i are constant matrices with proper dimensions. $i:[0,\infty) \to M = \{1,2,3...,m\}$ is the switching signal.

Lemma 1: X, Y are matrices with proper dimensions. There exists a scalar $\alpha > 0$ such that the following inequality holds：

$$X^T Y + Y^T X \le \alpha X^T X + \alpha^{-1} Y^T Y \tag{3}$$

Lemma 2: For given symmetric matrix $S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$, the dimension of S_{11} is $r \times r$. The following three conditions are equivalent：

$$\begin{aligned} &(1) \quad S < 0 \\ &(2) \quad S_{11} < 0, S_{22} - S_{12}^T S_{11}^{-1} S_{12} < 0 \\ &(3) \quad S_{22} < 0, S_{11} - S_{12} S_{22}^{-1} S_{12}^T < 0 \end{aligned} \tag{4}$$

Lemma 3: For a given scalar $\gamma > 0$, $\gamma = \max(\gamma_i), i \in M$, if there is a switching law $i = i(x(t),t)$ and a positive matrix P_i satisfying

$$\begin{bmatrix} A_i^T P_i + P_i A_i & P_i B_i & C_i^T \\ * & -\gamma_i^{-2} I & 0 \\ * & * & -I \end{bmatrix} < 0 \tag{5}$$

the system (2) is stable with $H\infty$ performance γ.

Proof: Choose the Lyapunov function of the sub-system of system (2) as $V_i = x^T P_i x$. The derivative of the Lyapunov function is

$$\dot{V}_i = \dot{x}^T P_i x + x^T P_i \dot{x} = x^T (A_i^T P_i + P_i A_i) x + w^T B_i^T P_i x + x^T B_i P_i w$$

When w=0, if the above equation satisfies

$$x^T (A_i^T P_i + P_i A_i) x < 0$$

the system (2) is asymptotic stable.

According Lemma 2, the inequality (5) is equivalent to

$$\begin{bmatrix} A_i^T P_i + P_i A_i & P_i B_i \\ * & -\gamma_i^2 I \end{bmatrix} + \begin{bmatrix} C_i^T \\ 0 \end{bmatrix} \begin{bmatrix} C_i & 0 \end{bmatrix} < 0$$

For any x(t) and w(t), the following inequality holds

$$\begin{bmatrix} x^T(t) \\ w^T(t) \end{bmatrix} \left\{ \begin{bmatrix} A_i^T P_i + P_i A_i & P_i B_i \\ * & -\gamma_i^2 I \end{bmatrix} + \begin{bmatrix} C_i^T \\ 0 \end{bmatrix} \begin{bmatrix} C_i & 0 \end{bmatrix} \right\} \begin{bmatrix} x^T(t) \\ w^T(t) \end{bmatrix}^T < 0$$

which is equivalent to

$$\dot{V}_i(x(t), w(t), t) + z^T(t) z(t) - \gamma_i^2 w^T(t) w(t) < 0$$

Under the zero initial condition, we have

$$\begin{aligned} &\int_0^{+\infty} [z^T(t) z(t) - \gamma^2 w^T(t) w(t)] dt \\ &\le \int_0^{+\infty} [\dot{V}_i(x(t), w(t), t) + z^T(t) z(t) - \gamma_i^2 w^T(t) w(t)] dt - V_M(x(+\infty), w(+\infty), +\infty) \\ &= \sum_{i=1}^{M} \int_{t_{i-1}}^{t_i} \left[\dot{V}_i(x(t), w(t), t) + z^T(t) z(t) - \gamma_i^2 w^T(t) w(t) \right] dt - V_M(x(+\infty), w(+\infty), +\infty) < 0 \end{aligned}$$

Thus

$$\int_0^{\infty} z^T(t)z(t)dt \le \gamma^2 \int_0^{+\infty} w^T(t)w(t)dt$$

This completes the proof.

From the above proof we know that if inequality (5) holds,

1) When the disturbance w=0, the system is asymptotically stable;
2) There exist a scalar $\gamma > 0$ satisfying the robust H_∞ performance

$$\int_0^{\infty} z^T(t)z(t)dt \le \gamma^2 \int_0^{+\infty} w^T(t)w(t)dt$$

Therefore, we can conclude that the switched system (2) satisfies the condition of robust H_∞ control.

Consider the following linear switched system：

$$\begin{aligned} \dot{x} &= A_i x + B_i u + D_i w \\ z &= C_i x \end{aligned} \tag{6}$$

where, $x(t) \in R^n$ is the state vector, $u(t) \in R^{m1}$ is the control input, $z(t) \in R^{m2}$ is the output, $w(t) \in l_2$ is the disturbance. A_i, B_i, C_i, D_i are constant matrices with proper dimensions. $i:[0,\infty) \to M = \{1,2,3...,m\}$ is the switching signal.

Definition 1. For a given scalar $\gamma > 0$, $\gamma = \max(\gamma_i), i \in M$, if there is a state feedback control without memory $u = K_i x$, such that the closed-loop subsystem of system (6) is stable with H_∞ performance γ, the system (6) is robust stabilizable with H_∞ performance γ.

With the above knowledge, we will study linear switched systems with time delay in the following sections. Firstly, the robust H_∞ control for general linear switched systems is analyzed. The results are then extended to uncertain switched systems with time-varying delay and uncertain switched systems with multiple time delays.

3. Robust H_∞ stabilization for linear switched systems with time delay

Consider the following Linear switched systems with time delay

$$\begin{cases} \dot{x}(t) = A_{\sigma(t)}x(t) + A_{d\sigma(t)}x(t-\tau) + B_{\sigma(t)}u(t) + B_{w\sigma(t)}w(t) \\ z(t) = C_{\sigma(t)}x(t) \\ x(t) = \phi(t) \end{cases} \tag{7}$$

Where $x(t) \in R^n$ is the state vector, $u(t) \in R^{m1}$ is the control input, $z(t) \in R^{m2}$ is the output vector, $w(t) \in l_2$ is the disturbance, $z \in R^q$ is the controlled output, $\sigma(t):[0,\infty) \to M = \{1,2,3...,m\}$ is the switching signal, $A_{\sigma(t)}$, $B_{\sigma(t)}$, $A_{d\sigma(t)}$, $B_{w\sigma(t)}$, $C_{\sigma(t)}$ are known constant matrices, τ is the time delay, $\phi(t)$ is a smooth function on R^n presenting the initial condition of the system.

Theorem 1. For system (7), and given scalar $\gamma > 0$, $\gamma = \max(\gamma_i)$, If there exist a switching law $\sigma(t) = i$ and positive matrices $P_i, R_i \in R^{n \times n}$ such that the following inequality holds:

$$\begin{bmatrix} S_1 & P_iA_d & P_iB_w & C^T \\ * & -R_i & 0 & 0 \\ * & * & -\gamma_i^2 I & 0 \\ * & * & * & -I \end{bmatrix} < 0 \tag{8}$$

where $S_1 = (A+BK_i)^T P_i + P_i(A+BK_i) + R_i$, system (7) is stabilizable with $H\infty$ performance γ. The controller is $u(t) = K_i x(t)$ and the switching law is $\sigma(t) = \arg\min_{i\in M}\{x^T(t)P_i x(t)\}$。

Proof: Suppose there are positive definite matrices $P_i, R_i \in R^{n\times n}$ and matrix $K_i \in R^{m1\times n}$, such that the linear matrix inequality (8) holds. The controller is $u(t) = K_i x(t)$ and choose the Lyapunov function as

$$V(x(t), w(t), t) = x^T(t)P_i x(t) + \int_{t-\tau}^{t} x^T(s)R_i x(s)ds$$

Then,

$$\begin{aligned} \dot{V}(x(t), w(t), t) &= \dot{x}^T(t)P_i x(t) + x^T(t)P_i\dot{x}(t) + x^T(t)R_i x(t) - x^T(t-\tau)R_i x(t-\tau) \\ &= x^T(t)[(A+BK_i)^T P_i + P_i(A+BK_i) + R_i]x(t) + \\ &\quad x^T(t)P_iB_w w(t) + w^T(t)B_w^T P_i x(t) + x^T(t-\tau)A_d^T P_i x(t) \\ &\quad + x^T(t)P_iA_d x(t-\tau) - x^T(t-\tau)R_i x(t-\tau) \end{aligned} \tag{9}$$

When $w(t) = 0$, if (10) holds, the closed-loop system of system (7) is asymptotically stable. Thus,

$$\begin{aligned} &x^T(t)[(A+BK_i)^T P_i + P_i(A+BK_i) + R_i]x(t) + x^T(t-\tau)A_d^T P_i x(t) \\ &+ x^T(t)P_iA_d x(t-\tau) - x^T(t-\tau)R_i x(t-\tau) < 0 \end{aligned} \tag{10}$$

Rewrite the inequality (10) as

$$\begin{bmatrix} x(t) \\ x(t-\tau) \end{bmatrix}^T \begin{bmatrix} S_1 & P_iA_d \\ * & -R_i \end{bmatrix} \begin{bmatrix} x(t) \\ x(t-\tau) \end{bmatrix} < 0 \tag{11}$$

By Lemma 2 and inequality (8), we have

$$\begin{bmatrix} S_1 & P_iA_d & P_iB_w \\ * & -R_i & 0 \\ * & * & -\gamma_i^2 I \end{bmatrix} + \begin{bmatrix} C^T \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} C^T \\ 0 \\ 0 \end{bmatrix}^T < 0 \tag{12}$$

For any $x(t), x(t-\tau), w(t)$, the following inequality holds.

$$\begin{bmatrix} x(t) \\ x(t-\tau) \\ w(t) \end{bmatrix}^T \begin{bmatrix} S_1 & P_iA_d & P_iB_w \\ * & -R_i & 0 \\ * & * & -\gamma_i^2 I \end{bmatrix} \begin{bmatrix} x(t) \\ x(t-\tau) \\ w(t) \end{bmatrix} + \begin{bmatrix} x(t) \\ x(t-\tau) \\ w(t) \end{bmatrix}^T \begin{bmatrix} C^T \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} C^T \\ 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} x(t) \\ x(t-\tau) \\ w(t) \end{bmatrix} < 0 \tag{13}$$

Thus, we have

$$\dot{V}(x(t), w(t), t) + z^T(t)z(t) - \gamma_i^2 w^T(t)w(t) < 0$$

With $\gamma = \max(\gamma_i), i \in M$ and under zero initial condition, we have

$$\int_0^{+\infty}[z^T(t)z(t)-\gamma^2 w^T(t)w(t)]dt$$
$$\le \int_0^{+\infty}[\dot{V}(x(t),w(t),t)+z^T(t)z(t)-\gamma_i^2 w^T(t)w(t)]dt - V(x(+\infty),w(+\infty),+\infty)$$
$$=\sum_{i=1}^{M}\int_{t_{i-1}}^{t_i}[\dot{V}(x(t),w(t),t)+z^T(t)z(t)-\gamma_i^2 w^T(t)w(t)]dt - V(x(+\infty),w(+\infty),+\infty)<0$$

Thus

$$\int_0^{\infty} z^T(t)z(t)dt \le \gamma^2\int_0^{+\infty} w^T(t)w(t)dt$$

that is

$$z^T(t)z(t)dt < \gamma^2 w^T(t)w(t)\text{。}$$

This complete the proof.

Remark 1. To convert the inequality (8) into an LMI, right and left multiplying the following matrix

$$\begin{bmatrix} X_i & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix}$$

to inequality (8) lead to

$$\begin{bmatrix} S & A_d & B_w & X_i C^T \\ * & -R_i & 0 & 0 \\ * & * & -\gamma_i^2 I & 0 \\ * & * & * & -I \end{bmatrix} < 0 \tag{14}$$

where, $S = AX_i + X_i A^T + BY_i + Y_i^T B^T + Q_i$, $X_i = P_i^{-1}, Y_i = K_i X_i, Q_i = X_i R_i X_i$。 The state feedback gain $K_i = Y_i X_i^{-1}, i \in M$ can be obtained by solving inequality (14).

Example 1 Consider the linear switched system (7) with

$$\begin{aligned} &A^1 = \begin{bmatrix} -1 & -2 \\ 1 & 2 \end{bmatrix}, A_1^1 = \begin{bmatrix} -1 & -1 \\ 0 & 0 \end{bmatrix}, B^1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, B_1^1 = \begin{bmatrix} 0.2 \\ 0.5 \end{bmatrix}, C^1 = \begin{bmatrix} 1 & 1 \end{bmatrix}, \\ &\gamma^1 = 3, w^1(t) = \cos(t), \tau^1 = 0.2. \\ &A^2 = \begin{bmatrix} -1.5 & -2 \\ 1 & 3 \end{bmatrix}, A_1^2 = \begin{bmatrix} -1 & -1 \\ 0 & 0 \end{bmatrix}, B^2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, B_1^2 = \begin{bmatrix} 0.2 \\ 0.5 \end{bmatrix}, C^2 = \begin{bmatrix} 1 & 1 \end{bmatrix}, \\ &\gamma^2 = 3, w^2(t) = \cos(t), \tau^2 = 0.2. \end{aligned} \tag{15}$$

By solving the linear matrix inequality (14), we have:

$$P^1 = \begin{bmatrix} 0.1967 & -1.2115 \\ -1.2115 & 7.4606 \end{bmatrix}, K^1 = \begin{bmatrix} 1.1140 & -6.8599 \end{bmatrix}$$
$$P^2 = \begin{bmatrix} 0.0281 & -0.1732 \\ -0.1732 & 1.0670 \end{bmatrix}, K^2 = \begin{bmatrix} 0.7320 & -4.5086 \end{bmatrix}$$

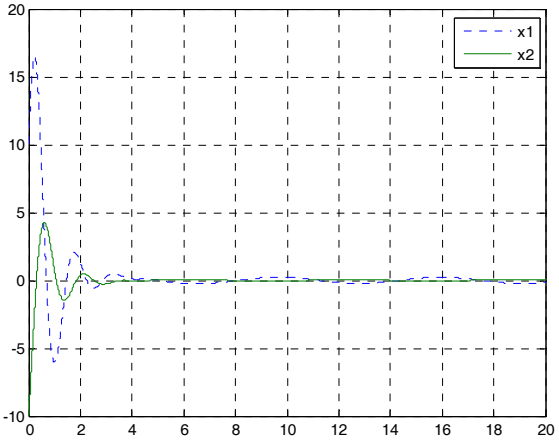


Fig. 1. State response of Example 1

By Theorem 1, the switching law is designed as:

$$\sigma(t) = \mathrm{i} = \begin{cases} 1, x^T(P^1 - P^2)x \le 0 \\ 2, x^T(P^1 - P^2)x > 0 \end{cases}$$

The state response of the closed-loop system is shown in Figure 1. x1 and x2 are system states, the initial condition is [x1,x2]=[10,-10]. The result shows the system is stable under the switching law when it is switched among the closed-loop subsystems.

4. Robust H_∞ stabilization for uncertain linear switched systems with time-varying delay

Consider the following linear uncertain switched system with time-varying delay

$$\begin{aligned} \dot{x}(t) &= (A_{\sigma(t)} + \Delta A_{\sigma(t)})x(t) + (A_{d\sigma(t)} + \Delta A_{d\sigma(t)})x(t - d(t)) + (B_{\sigma(t)} + \Delta B_{\sigma(t)})u(t) \\ &\quad + (B_{h\sigma(t)} + \Delta B_{h\sigma(t)})u(t - h(t)) + (B_{1w\sigma(t)} + \Delta B_{1w\sigma(t)})w(t) \\ z(t) &= (C_{\sigma(t)} + \Delta C_{\sigma(t)})x(t) + (C_{d\sigma(t)} + \Delta C_{d\sigma(t)})x(t - d(t)) + (D_{\sigma(t)} + \Delta D_{\sigma(t)})u(t) \\ &\quad + (D_{h\sigma(t)} + \Delta D_{h\sigma(t)})u(t - h(t)) + (B_{2w\sigma(t)} + \Delta B_{2w\sigma(t)})w(t) \\ x(t) &= \phi(t), t \in [-\max(d(t), h(t)), 0] \end{aligned} \tag{16}$$

where , $x(t) \in R^n$ is state vector, $u(t) \in R^{m1}$ is the control input, $z(t) \in R^{m2}$ is the output, $w(t) \in l_2$ is the disturbance, $\sigma(t) : [0, \infty) \to M = \{1, 2, 3..., m\}$ is the switching signal, $A_{\sigma(t)}$, $A_{d\sigma(t)}$, $B_{\sigma(t)}$, $B_{h\sigma(t)}$, $B_{1w\sigma(t)}$, $C_{\sigma(t)}$, $C_{d\sigma(t)}$, $D_{\sigma(t)}$, $D_{h\sigma(t)}$, $B_{2w\sigma(t)}$ are known constant matrices, $\Delta A_{d\sigma(t)}$, $\Delta B_{\sigma(t)}$, $\Delta B_{h\sigma(t)}$, $\Delta B_{1w\sigma(t)}$, $\Delta C_{\sigma(t)}$, $\Delta C_{d\sigma(t)}$, $\Delta D_{\sigma(t)}$, $\Delta D_{h\sigma(t)}$, $\Delta B_{2w\sigma(t)}$ are bounded real functional matrices with proper dimensions, representing the uncertainties, $\phi(t)$ is the initial condition. d(t) and h(t) are the state delay and the control delay respectively. There are positive scalers d, h, ρ_d, ρ_h, such that

$$\begin{aligned} &0 \le d(t) \le d < \infty, \dot{d}(t) \le \rho_d < 1 \\ &0 \le h(t) \le h < \infty, \dot{h}(t) \le \rho_h < 1 \end{aligned} \tag{17}$$

Denote

$$\begin{aligned} &\overline{A} = A_{\sigma(t)} + \Delta A_{\sigma(t)}, \overline{A}_d = A_{d\sigma(t)} + \Delta A_{d\sigma(t)}, \overline{B} = B_{\sigma(t)} + \Delta B_{\sigma(t)} \\ &\overline{B}_h = B_{h\sigma(t)} + \Delta B_{h\sigma(t)}, \overline{B}_1 = B_{1w\sigma(t)} + \Delta B_{1w\sigma(t)} \\ &\overline{C} = C_{\sigma(t)} + \Delta C_{\sigma(t)}, \overline{C}_d = C_{d\sigma(t)} + \Delta C_{d\sigma(t)}, \overline{D} = D_{\sigma(t)} + \Delta D_{\sigma(t)} \\ &\overline{D}_h = D_{h\sigma(t)} + \Delta D_{h\sigma(t)}, \overline{B}_2 = B_{2w\sigma(t)} + \Delta B_{2w\sigma(t)} \end{aligned} \tag{18}$$

and suppose:

$$\begin{bmatrix} \Delta A_{\sigma(t)} & \Delta A_{d\sigma(t)} & \Delta B_{\sigma(t)} \Delta B_{h\sigma(t)} & \Delta B_{1w\sigma(t)} \\ \Delta C_{\sigma(t)} & \Delta C_{d\sigma(t)} & \Delta D_{\sigma(t)} \Delta D_{h\sigma(t)} & \Delta B_{2w\sigma(t)} \end{bmatrix} = \begin{bmatrix} H_{1\sigma} \\ H_{2\sigma} \end{bmatrix} F_{\sigma(t)} \begin{bmatrix} E_{1\sigma} & E_{2\sigma} & E_{3\sigma} & E_{4\sigma} & E_{5\sigma} \end{bmatrix} \tag{19}$$

where $H_{1\sigma}, H_{2\sigma}, E_{1\sigma}, E_{2\sigma}, E_{3\sigma}, E_{4\sigma}, E_{5\sigma}$ are constant matrices with proper dimensions and $F_{\sigma(t)}$ satisfying

$$F_{\sigma(t)}^T F_{\sigma(t)} \le I \tag{20}$$

Theorem 2 For a given scalar $\gamma_i > 0$, if there are positive definite matrices $P_i, R_{1i}, R_{2i} \in R^{n\times n}$, such that：

$$\begin{bmatrix} S_1 & P_i\overline{A}_d & P_i\overline{B}_h K_i & P_i\overline{B}_1 & \overline{C}^T + K_i^T \overline{D}^T \\ * & -(1-\rho_d)R_{1i} & 0 & 0 & \overline{C}_d^T \\ * & * & -(1-\rho_h)R_{2i} & 0 & K_i^T \overline{D}_h^T \\ * & * & * & -\gamma_i^2 I & \overline{B}_2^T \\ * & * & * & * & -I \end{bmatrix} < 0 \tag{21}$$

where

$$S_1 = \overline{A}^T P_i + P_i \overline{A} + K_i^T \overline{B}^T P_i + P_i \overline{B} K_i + R_{1i} + R_{2i}$$

the system (16) is robust stabilizable with $H\infty$ performance γ, $\gamma = \max(\gamma_i)$. $u(t) = K_i x(t)$ is the switched robust $H\infty$ controller. The switching law is $\sigma(t) = i = \arg\min_{i\in M}\{x^T(t)P_i x(t)\}$。

Proof: If there are positive definite matrices $P_i, R_{1i}, R_{2i} \in R^{n\times n}$ and matrix $K_i \in R^{m1\times n}$ satisfying the inequality (21) with the controller $u(t) = K_i x(t)$ and Lyapunov function:

$$V(x(t), w(t), t) = x^T(t)P_i x(t) + \int_{t-d(t)}^{t} x^T(s)R_{1i}x(s)ds + \int_{t-h(t)}^{t} x^T(s)R_{2i}x(s)ds$$

Then

$$\begin{aligned} \dot{V}(x(t), w(t), t) &= \dot{x}^T(t)P_i x(t) + x^T(t)P_i\dot{x}(t) + x^T(t)(R_{1i} + R_{2i})x(t) \\ &\quad -(1-\dot{d}(t))x^T(t-d(t))R_{1i}x(t-d(t)) - (1-\dot{h}(t))x^T(t-h(t))R_{2i}x(t-h(t)) \\ &= x^T(t)\overline{A}^T P_i x(t) + x^T(t)P_i\overline{A}x(t) + x^T(t)K_i^T\overline{B}^T P_i x(t) + x^T(t)P_i\overline{B}K_i x(t) \\ &\quad + x^T(t)(R_{1i} + R_{2i})x(t) + x^T(t)P_i\overline{B}_1 w(t) + w^T(t)\overline{B}_1^T P x(t) \\ &\quad + x^T(t-d(t))\overline{A}_d^T P_i x(t) + x^T(t)P_i\overline{A}_d x(t-d(t)) + x^T(t-h(t))K_i^T\overline{B}_h^T P_i x(t) \\ &\quad + x^T(t)P_i\overline{B}_h K_i x(t-h(t)) - (1-\dot{d}(t))x^T(t-d(t))R_{1i}x(t-d(t)) \\ &\quad -(1-\dot{h}(t))x^T(t-h(t))R_{2i}x(t-h(t)) \end{aligned} \tag{22}$$

When $w(t) = 0$, considering condition (17), if inequality (23) holds, the closed-loop system is robust asymptotically stable

$$\begin{aligned}
&x^T(t)\bar{A}^T P_i x(t) + x^T(t) P_i \bar{A} x(t) + x^T(t) K_i^T \bar{B}^T P_i x(t) + x^T(t) P_i \bar{B} K_i x(t) \\
&+x^T(t)(R_{1i}+R_{2i})x(t) + x^T(t-d(t))\bar{A}_d^T P_i x(t) + x^T(t) P_i \bar{A}_d x(t-d(t)) \\
&+x^T(t-h(t))K_i^T \bar{B}_h^T P_i x(t) + x^T(t) P_i \bar{B}_h K_i x(t-h(t)) \\
&-(1-\rho_d)x^T(t-d(t))R_{1i}x(t-d(t)) - (1-\rho_h)x^T(t-h(t))R_{2i}x(t-h(t)) < 0
\end{aligned} \tag{23}$$

Rewrite inequality (23) as

$$\begin{bmatrix} x^T & x^T(t-d(t)) & x^T(t-h(t)) \end{bmatrix} W \begin{bmatrix} x^T & x^T(t-d(t)) & x^T(t-h(t)) \end{bmatrix}^T < 0 \tag{24}$$

where

$$W = \begin{bmatrix} \bar{A}^T P_i + P_i \bar{A} + K_i^T \bar{B} K_i + R_{1i} + R_{2i} & P_i \bar{A}_d & P_i \bar{B}_h K_i \\ * & -(1-\rho_d)R_{1i} & 0 \\ * & * & -(1-\rho_h)R_{2i} \end{bmatrix} < 0$$

By inequality (21) and Lemma 2, the following inequality follows

$$\begin{aligned}
&\begin{bmatrix} \bar{A}^T P_i + P_i \bar{A} + K_i^T \bar{B} K + R_{1i} + R_{2i} & P_i \bar{A}_d & P_i \bar{B}_h K_i & P_i \bar{B}_1 \\ * & -(1-\rho_d)R_{1i} & 0 & 0 \\ * & * & -(1-\rho_h)R_{2i} & 0 \\ * & * & * & -\gamma_i^2 I \end{bmatrix} + \\
&\begin{bmatrix} \bar{C}^T + K_i^T \bar{D}^T \\ \bar{C}_d^T \\ K_i^T \bar{D}_h^T \\ \bar{B}_3^T \end{bmatrix} \begin{bmatrix} \bar{C} + \bar{D} K_i & \bar{C}_d & \bar{D}_h K_i & \bar{B}_2 \end{bmatrix} < 0
\end{aligned} \tag{25}$$

For any $x(t), x(t-d(t)), x(t-h(t)), w(t)$, the following inequality holds

$$\begin{aligned}
&\begin{bmatrix} x(t) \\ x(t-d(t)) \\ x(t-h(t)) \\ w(t) \end{bmatrix}^T \begin{bmatrix} \bar{A}^T P_i + P_i \bar{A} + K_i^T \bar{B} K + R_{1i} + R_{2i} & P_i \bar{A}_d & P_i \bar{B} K_i & P_i \bar{B}_1 \\ * & -(1-\rho_d)R_{1i} & 0 & 0 \\ * & * & -(1-\rho_h)R_{2i} & 0 \\ * & * & * & -\gamma_i^2 I \end{bmatrix} \begin{bmatrix} x(t) \\ x(t-d(t)) \\ x(t-h(t)) \\ w(t) \end{bmatrix} \\
&+ \begin{bmatrix} x(t) \\ x(t-d(t)) \\ x(t-h(t)) \\ w(t) \end{bmatrix}^T \begin{bmatrix} \bar{C}^T + K_i^T \bar{D}^T \\ \bar{C}_d^T \\ K_i^T \bar{D}_h^T \\ \bar{B}_2^T \end{bmatrix} \begin{bmatrix} \bar{C} + \bar{D} K_i & \bar{C}_d & \bar{D}_h K_i & \bar{B}_2 \end{bmatrix} \begin{bmatrix} x(t) \\ x(t-d(t)) \\ x(t-h(t)) \\ w(t) \end{bmatrix} < 0
\end{aligned} \tag{26}$$

Thus

$$\dot{V} + z^T(t) z(t) - \gamma^2 w^T(t) w(t) < 0 \tag{27}$$

Under zero initial condition with $\gamma = \max(\gamma_i), i \in M$, we have

$$\begin{aligned}
&\int_0^{+\infty} [z^T(t) z(t) - \gamma^2 w^T(t) w(t)] dt \\
&\le \int_0^{+\infty} [\dot{V}(x(t), w(t), t) + z^T(t) z(t) - \gamma_i^2 w^T(t) w(t)] dt - V(x(+\infty), w(+\infty), +\infty) \\
&= \sum_{i=1}^{M} \int_{t_{i-1}}^{t_i} [\dot{V}(x(t), w(t), t) + z^T(t) z(t) - \gamma_i^2 w^T(t) w(t)] dt - V(x(+\infty), w(+\infty), +\infty) < 0
\end{aligned} \tag{28}$$

Therefore

$$\int_0^{\infty} z^T(t)z(t)dt \le \gamma^2 \int_0^{+\infty} w^T(t)w(t)dt$$

This completes the proof.

Remark 2: Although theorem 2 presents a sufficient condition for the robust stabilization with $H\infty$ performance γ, there are still uncertainties in the inequality (21).

Theorem 3 For the switched system (16) and a given positive scalar γ_i, if there are matrix Y_i with proper dimension, positive definite matrices X_i, Q_{1i}, Q_{2i} and scalar $\alpha > 0$ such that

$$\begin{bmatrix} S & A_d X_i & B_h Y_i & B_{1w} & \varphi & X_i E_1^T + Y_i^T E_3^T \\ * & -(1-\rho_d)Q_{1i} & 0 & 0 & X_i C_d^T & X_i E_2^T \\ * & * & -(1-\rho_h)Q_{2i} & 0 & Y_i^T D_h^T & Y_i^T E_4^T \\ * & * & * & -\gamma_i^2 I & B_{2w}^T & E_5^T \\ * & * & * & * & -I + \alpha H_2 H_2^T & 0 \\ * & * & * & * & * & -\alpha I \end{bmatrix} < 0 \tag{29}$$

where

$$S = X_i A^T + A X_i + Y_i^T B^T + B Y_i + Q_{1i} + Q_{2i} + \alpha H_1 H_1^T,$$

and

$$\varphi = X_i C^T + Y_i^T D^T + \alpha H_1 H_2^T$$

holds, the system is robust stabilizable with $H\infty$ performance γ, $\gamma = \max(\gamma_i)$. The robust $H\infty$ switched controller is given by:

$$K_i = Y_i X_i^{-1}, i \in M \tag{30}$$

And the switching law is

$$\sigma(t) = i = \arg\min_{i \in M} \{x^T(t) X_i^{-1} x(t)\} \tag{31}$$

Proof: For any non-zero vector ξ, by inequality (21), we have

$$\xi^T \begin{bmatrix} S_1 & P_i \bar{A}_d & P_i \bar{B}_h K_i & P_i \bar{B}_1 & \bar{C}^T + K_i^T \bar{D}^T \\ * & -(1-\rho_d)R_{1i} & 0 & 0 & \bar{C}_d^T \\ * & * & -(1-\rho_h)R_{2i} & 0 & K_i^T \bar{D}_h^T \\ * & * & * & -\gamma_i^2 I & \bar{B}_2^T \\ * & * & * & * & -I \end{bmatrix} \xi = \xi^T L \xi + \xi^T \Delta L \xi < 0 \tag{32}$$

where

$$L = \begin{bmatrix} S_2 & P_i A_d & P_i B_h K_i & P_i B_{1w} & C^T + K_i^T D^T \\ * & -(1-\rho_d)R_{1i} & 0 & 0 & C_d^T \\ * & * & -(1-\rho_h)R_{2i} & 0 & K_i^T D_h^T \\ * & * & * & -\gamma_i^2 I & B_{2w}^T \\ * & * & * & * & -I \end{bmatrix} \tag{33}$$

$$S_2 = A^T P_i + P_i A + K_i B^T P_i + P_i B K_i + R_{1i} + R_{2i}$$

and

$$\Delta L=\begin{bmatrix} P_iH_1 \\ 0 \\ 0 \\ 0 \\ H_2 \end{bmatrix} F(t)\begin{bmatrix} E_1+E_3K_i & E_2 & E_4K_i & E_5 & 0 \end{bmatrix}+\begin{bmatrix} E_1+E_3K_i & E_2 & E_4K_i & E_5 & 0 \end{bmatrix}^T F^T(t)\begin{bmatrix} P_i \\ 0 \\ 0 \\ 0 \\ H_2 \end{bmatrix}^T \tag{34}$$

By Lemma 1 and inequality (20), we have

$$\begin{aligned} \Delta L=&\begin{bmatrix} P_iH_1 \\ 0 \\ 0 \\ 0 \\ H_2 \end{bmatrix} F(t)F^T(t)\begin{bmatrix} P_iH_1 \\ 0 \\ 0 \\ 0 \\ H_2 \end{bmatrix}^T \\ &+\begin{bmatrix} E_1+E_3K_i & E_2 & E_4K_i & E_5 & 0 \end{bmatrix}^T F(t)F^T(t)\begin{bmatrix} E_1+E_3K_i & E_2 & E_4K_i & E_5 & 0 \end{bmatrix} \\ &\le \alpha L_1+\alpha^{-1}L_2 \end{aligned} \tag{35}$$

where $\alpha>0$,

$$L_1=\begin{bmatrix} P_iH_1 \\ 0 \\ 0 \\ 0 \\ H_2 \end{bmatrix}\begin{bmatrix} P_iH_1 \\ 0 \\ 0 \\ 0 \\ H_2 \end{bmatrix}^T$$

$$L_2=\begin{bmatrix} E_1+E_3K_i & E_2 & E_4K_i & E_5 & 0 \end{bmatrix}^T\begin{bmatrix} E_1+E_3K_i & E_2 & E_4K_i & E_5 & 0 \end{bmatrix}$$

If

$$L+\alpha L_1+\alpha^{-1}L_2<0 \tag{36}$$

the inequality (21) follows

By Lemma 2, inequality (36) is equivalent to:

$$\begin{bmatrix} S_3 & P_iA_d & P_iB_hK_i & P_iB_{1w} & C^T+K_i^TD^T+\alpha P_iH_1H_2^T & E_1^T+K_i^TE_3^T \\ * & -(1-\rho_d)R_{1i} & 0 & 0 & C_d^T & E_2^T \\ * & * & -(1-\rho_h)R_{2i} & 0 & K_i^TD_h^T & K_i^TE_4^T \\ * & * & * & -\gamma_i^2I & B_{2w}^T & E_5^T \\ * & * & * & * & -I+\alpha H_2H_2^T & 0 \\ * & * & * & * & * & -\alpha I \end{bmatrix}<0 \tag{37}$$

Denote

$$X_i=P_i^{-1},Q_{1i}=X_iR_{1i}X_i,Q_{2i}=X_iR_{2i}X_i,$$

and

$$K_i=Y_iX_i^{-1}$$

where $S_3=A^TP_i+P_iA+K_i^TB^TP_i+P_iBK_i+R_{1i}+R_{2i}+\alpha P_iH_1H_1^TP_i$, Y_i is an arbitrary matrix with proper dimension. Q_{1i},Q_{2i} are positive definite matrices with proper dimensions. By right and left multiplying the following matrix to the inequality (37)

$$\begin{bmatrix} X_i & 0 & 0 & 0 & 0 & 0 \\ 0 & X_i & 0 & 0 & 0 & 0 \\ 0 & 0 & X_i & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & I \end{bmatrix}$$

the inequality (29) follows.
This completes the proof.
Example 2 Consider uncertain switched system (16) with

$$A^1 = \begin{bmatrix} -2 & 3 \\ 3 & 4 \end{bmatrix}, A1^1 = \begin{bmatrix} -2 & 1 \\ 0 & 0 \end{bmatrix}, B^1 = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}, B1^1 = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.3 \end{bmatrix},$$
$$B2^1 = \begin{bmatrix} 0.5 \\ 0.3 \end{bmatrix}, B3^1 = \begin{bmatrix} 0.5 \\ 0.4 \end{bmatrix}, C^1 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, C1^1 = \begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 0.1 \end{bmatrix}, D^1 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.4 \end{bmatrix},$$
$$D1^1 = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.2 \end{bmatrix}, H1^1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, H2^1 = \begin{bmatrix} 0.5 & 0.4 \\ 0 & 0 \end{bmatrix}, E1^1 = \begin{bmatrix} 0.5 & 0.3 \\ 0 & 0.6 \end{bmatrix},$$

$$E2^1 = \begin{bmatrix} 0.5 & 0.4 \\ 0 & 0 \end{bmatrix}, E3^1 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}, E4^1 = \begin{bmatrix} 0 & 0 \\ 0 & 0.5 \end{bmatrix}, E5^1 = \begin{bmatrix} 0.2 \\ 0.2 \end{bmatrix},$$
$$A^2 = \begin{bmatrix} -1 & 1 \\ 3 & 2 \end{bmatrix}, A1^2 = \begin{bmatrix} -1 & -1 \\ 0 & 2 \end{bmatrix}, B^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, B1^2 = \begin{bmatrix} 0.2 & 0 \\ 0 & 0 \end{bmatrix},$$
$$B2^2 = \begin{bmatrix} 0.5 \\ 0.3 \end{bmatrix}, B3^2 = \begin{bmatrix} 0.5 \\ 0.4 \end{bmatrix}, C^2 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, C1^2 = \begin{bmatrix} 0.3 & 0.5 \\ 0.2 & 0.1 \end{bmatrix}, D^2 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.1 \end{bmatrix},$$
$$D1^2 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.2 \end{bmatrix}, H1^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, H2^2 = \begin{bmatrix} 0.5 & 0.4 \\ 0 & 0 \end{bmatrix}, E1^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$
$$E2^2 = \begin{bmatrix} 0.5 & 0.4 \\ 0 & 0 \end{bmatrix}, E3^2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, E4^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0.5 \end{bmatrix}, E5^2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (38)$$

Choose

$$F^1(t) = \sin(t), F^2(t) = \sin(t), w^1(t) = \mathrm{c1os}(t), w^2(t) = \cos(t), \alpha^1 = \alpha^2 = 0.5,$$
$$d^1(t) = 0.3\sin(t/3.85), d^2(t) = 0.3\sin(t/3.85), h^1(t) = 0.2\sin(t/3.85),$$
$$h^2(t) = 0.2\sin(t/3.85), \rho^1{}_d = 0.3, \rho^1{}_h = 0.2, \gamma_1 = 3, \rho^2{}_d = 0.3, \rho^2{}_h = 0.2, \gamma_2 = 3$$

By Theorem 3, we have:

$$P^1 = \begin{bmatrix} 4.0697 & 0 \\ 0 & 4.0697 \end{bmatrix}, K^1 = \begin{bmatrix} -8.6086 & -0.7780 \\ -2.3540 & -6.1658 \end{bmatrix}$$
$$P^2 = \begin{bmatrix} 1.9297 & 3.1929 \\ 3.1929 & 5.2830 \end{bmatrix}, K^2 = \begin{bmatrix} -0.6417 & -1.0618 \\ -3.6147 & -5.9809 \end{bmatrix}$$

and the switching law is designed as

$$\sigma(t) = \mathrm{i} = \begin{cases} 1, x^T(P^1 - P^2)x \le 0 \\ 2, x^T(P^1 - P^2)x > 0 \end{cases}$$

The state response is shown in Figure 2.

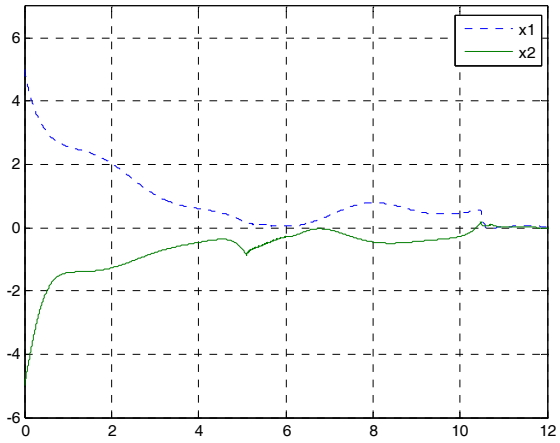


Fig. 2. State response of Example 2

x1 and x2 are system states. The initial condition is [x1,x2]=[5,-5]。The result shows the system is stable under the switching law when it is switched among the closed-loop subsystems.

5. State feedback robust H_∞ stabilization for linear uncertain switched systems with multiple time delays

Consider the following linear switched system with multiple time delays

$$\begin{aligned}
&\dot{x}(t)=(A_{\sigma(t)}+\Delta A_{\sigma(t)})x(t)+\sum_{j=1}^{N}(A_{dj\sigma(t)}+\Delta A_{dj\sigma(t)})x(t-\tau_j)\\
&+(B_{\sigma(t)}+\Delta B_{\sigma(t)})u(t)+B_{1\sigma(t)}w(t)\\
&z(t)=(C_{\sigma(t)}+\Delta C_{\sigma(t)})x(t)+\sum_{j=1}^{N}(C_{dj\sigma(t)}+\Delta C_{dj\sigma(t)})x(t-\tau_j)\\
&+(D_{\sigma(t)}+\Delta D_{\sigma(t)})u(t)+B_{2\sigma(t)}w(t)\\
&x(t)=\phi(t),t\in(-\max(\tau_j),0)
\end{aligned} \tag{39}$$

where $x(t)\in R^n$ is the system state vector, $u(t)\in R^{m1}$ is the control input of the system, $z(t)\in R^{m2}$ is the output vector, $w(t)\in l_2$ is the disturbance, $\sigma(t):[0,\infty)\to M=\{1,2,3...,m\}$ is the switching signal, $A_{\sigma(t)}$, $A_{dj\sigma(t)}$, $B_{\sigma(t)}$, $B_{1\sigma(t)}$, $C_{\sigma(t)}$, $C_{dj\sigma(t)}$, $D_{\sigma(t)}$, $B_{2\sigma(t)}$ are known constant matrices, $\Delta A_{d\sigma(t)}$, $\Delta A_{dj\sigma(t)}$, $\Delta B_{\sigma(t)}$, $\Delta C_{\sigma(t)}$, $\Delta C_{dj\sigma(t)}$, $\Delta D_{\sigma(t)}$ are bounded time-varying real functional matrices with proper dimensions, denoting the uncertainties of the switched systems. $\phi(t)$ is the initial condition of the system. τ_j is the delay of the system state.

Denote

$$\begin{aligned}&\bar{A}=A_{\sigma(t)}+\Delta A_{\sigma(t)},\bar{A}_{dj}=A_{dj\sigma(t)}+\Delta A_{dj\sigma(t)},\bar{B}=B_{\sigma(t)}+\Delta B_{\sigma(t)}\\&\bar{C}=C_{\sigma(t)}+\Delta C_{\sigma(t)},\bar{C}_{dj}=C_{dj\sigma(t)}+\Delta C_{dj\sigma(t)}\bar{D}=D_{\sigma(t)}+\Delta D_{\sigma(t)}\end{aligned}\tag{40}$$

and suppose：

$$\begin{bmatrix}\Delta A_{\sigma(t)} & \Delta A_{dj\sigma(t)} & \Delta B_{\sigma(t)}\\ \Delta C_{\sigma(t)} & \Delta C_{dj\sigma(t)} & \Delta D_{\sigma(t)}\end{bmatrix}=\begin{bmatrix}H_{1\sigma}\\ H_{2\sigma}\end{bmatrix}F_{\sigma(t)}\begin{bmatrix}E_{1\sigma} & E_{2j\sigma} & E_{3\sigma}\end{bmatrix}\tag{41}$$

where $H_{1\sigma},H_{2\sigma},E_{1\sigma},E_{2j\sigma},E_{3\sigma}$ are real constant matrices with proper dimensions, and $F_{\sigma(t)}$ satisfies:

$$F_{\sigma(t)}^{T}F_{\sigma(t)}\le I\tag{42}$$

Theorem 4. For a given scalar $\gamma_i>0$, if there are positive definite matrices $P_i,Q_{1i},...,Q_{Ni}\in R^{n\times n}$, such that:

$$\begin{bmatrix}S_1 & P_i(\bar{A}_{d1}+\bar{B}K_{1i}) & \dots & P_i(\bar{A}_{dN}+\bar{B}K_{Ni}) & P_iB_1 & \bar{C}^T+K_i^T\bar{D}^T\\ * & -Q_{1i} & 0 & 0 & 0 & \bar{C}_{d1}^T+K_{1i}^T\bar{D}^T\\ * & 0 & \dots & 0 & 0 & \dots\\ * & 0 & 0 & -Q_{Ni} & 0 & \bar{C}_{dN}^T+K_{Ni}^T\bar{D}^T\\ * & * & * & * & -\gamma_i^2I & B_2^T\\ * & * & * & * & * & -I\end{bmatrix}<0\tag{43}$$

where $S_1=(\bar{A}^T+K_i^T\bar{B}^T)P_i+P_i(\bar{A}+\bar{B}K_i)+\sum_{j=1}^{N}Q_{ji}$, the system (39) is robust stabilizable with H_∞ performance γ , $\gamma=\max(\gamma_i)$. The state feedback switched H_∞ control with memory is $u(t)=K_ix(t)+\sum_{j=1}^{N}K_{ji}x(t-\tau_{ji})$

Proof: Suppose there are positive definite matrices $P_i,Q_{1i},...,Q_{iN}\in R^{n\times n}$ and matrices $K_i,K_{ji}\in R^{m1\times n}$, such that the inequality (43) holds. The state feedback control is $u(t)=K_ix(t)+\sum_{j=1}^{N}K_{ji}x(t-\tau_{iu})$. Choose the Lyapunov function as

$$V(x(t),w(t),t)=x^T(t)P_ix(t)+\sum_{j=1}^{N}\int_{t-\tau_j}^{t}x^T(s)Q_{ji}x(s)ds$$

The derivative of the Lyapunov function is:

$$\begin{aligned}&\dot{V}(x(t),w(t),t)=\dot{x}^T(t)P_ix(t)+x^T(t)P_i\dot{x}(t)+\sum_{j=1}^{N}x^T(t)Q_{ji}x(t)-\sum_{j=1}^{N}x^T(t-\tau_j)Q_{ji}x(t-\tau_{ji})\\&=x^T(t)[P_i(\bar{A}+\bar{B}K_i)+(\bar{A}+\bar{B}K_i)^TP_i+\sum_{j=1}^{N}Q_{ji}]x(t)+\sum_{j=1}^{N}x^T(t-\tau_{ji})(\bar{A}_{dj}+\bar{B}K_{ji})^TP_ix(t)\\&+\sum_{j=1}^{N}x^T(t)P_i(\bar{A}_{dj}+\bar{B}K_{ji})x(t-\tau_{ji})+w^T(t)B_1^TP_ix(t)+x^T(t)P_iB_1w(t)-\sum_{j=1}^{N}x^T(t-\tau_{ji})Q_{ji}x(t-\tau_{ji})\end{aligned}\tag{44}$$

When $w(t)=0$, if the following inequality holds, the closed-loop is asymptotically stable

$$x^T(t)[P_i(\bar{A}+\bar{B}K_i)+(\bar{A}+\bar{B}K_i)^T P_i+\sum_{j=1}^{N}Q_{ji}]x(t)+\sum_{j=1}^{N}x^T(t-\tau_{ji})(\bar{A}_{dj}+\bar{B}K_{ji})^T P_i x(t)$$
$$+\sum_{j=1}^{N}x^T(t)P_i(\bar{A}_{dj}+\bar{B}K_{ji})x(t-\tau_{ji})-\sum_{j=1}^{N}x^T(t-\tau_{ji})Q_{ji}x(t-\tau_{ji})<0 \tag{45}$$

Rewrite the above inequality as

$$\begin{bmatrix} x(t) \\ x(t-\tau_{1i}) \\ ... \\ x(t-\tau_{Ni}) \end{bmatrix}^T \begin{bmatrix} S_1 & P_i(\bar{A}_{d1}+\bar{B}K_{1i}) & ... & P_i(\bar{A}_{dN}+\bar{B}K_{Ni}) \\ * & -Q_{1i} & 0 & 0 \\ * & 0 & ... & 0 \\ * & 0 & 0 & -Q_{Ni} \end{bmatrix} \begin{bmatrix} x(t) \\ x(t-\tau_{1i}) \\ ... \\ x(t-\tau_{Ni}) \end{bmatrix} < 0$$

By Lemma 2 and inequality (43), we have

$$\begin{bmatrix} S_1 & P_i(\bar{A}_{d1}+\bar{B}K_{1i}) & ... & P_i(\bar{A}_{dN}+\bar{B}K_{Ni}) & P_iB_1 \\ * & -Q_{1i} & 0 & 0 & 0 \\ * & * & ... & 0 & 0 \\ * & * & * & -Q_{Ni} & 0 \\ * & * & * & * & -\gamma_i^2 I \end{bmatrix} + \begin{bmatrix} \bar{C}^T+K_i^T\bar{D}^T \\ \bar{C}_{d1}^T+K_{1i}^T\bar{D}^T \\ ... \\ \bar{C}_{dN}^T+K_{Ni}^T\bar{D}^T \\ B_2^T \end{bmatrix} \begin{bmatrix} \bar{C}^T+K_i^T\bar{D}^T \\ \bar{C}_{d1}^T+K_{1i}^T\bar{D}^T \\ ... \\ \bar{C}_{dN}^T+K_{Ni}^T\bar{D}^T \\ B_2^T \end{bmatrix}^T < 0 \tag{46}$$

For any $x(t), x(t-\tau_{1i}), ...x(t-\tau_{Ni}), w(t)$, the following inequality holds.

$$\begin{bmatrix} x(t) \\ x(t-\tau_{1i}) \\ ... \\ x(t-\tau_{Ni}) \\ w(t) \end{bmatrix}^T \begin{bmatrix} S_1 & P_i(\bar{A}_{d1}+\bar{B}K_{1i}) & ... & P_i(\bar{A}_{dN}+\bar{B}K_{Ni}) & P_iB_1 \\ * & -Q_{i1} & 0 & 0 & 0 \\ * & * & ... & 0 & 0 \\ * & * & * & -Q_{iN} & 0 \\ * & * & * & * & -\gamma_i^2 I \end{bmatrix} \begin{bmatrix} x(t) \\ x(t-\tau_{1i}) \\ ... \\ x(t-\tau_{Ni}) \\ w(t) \end{bmatrix} +$$
$$\begin{bmatrix} x(t) \\ x(t-\tau_{1i}) \\ ... \\ x(t-\tau_{Ni}) \\ w(t) \end{bmatrix}^T \begin{bmatrix} \bar{C}^T+K_i^T\bar{D}^T \\ \bar{C}_{d1}^T+K_{1i}^T\bar{D}^T \\ ... \\ \bar{C}_{dN}^T+K_{Ni}^T\bar{D}^T \\ B_2^T \end{bmatrix} \begin{bmatrix} \bar{C}^T+K_i^T\bar{D}^T \\ \bar{C}_{d1}^T+K_{1i}^T\bar{D}^T \\ ... \\ \bar{C}_{dN}^T+K_{Ni}^T\bar{D}^T \\ B_2^T \end{bmatrix}^T \begin{bmatrix} x(t) \\ x(t-\tau_{1i}) \\ ... \\ x(t-\tau_{Ni}) \\ w(t) \end{bmatrix} < 0 \tag{47}$$

Thus

$$\dot{V}(x(t),w(t),t)+z^T(t)z(t)-\gamma_i^2 w^T(t)w(t)<0$$

Under the zero initial condition, by setting $\gamma=\max(\gamma_i), i\in M$, we have

$$\int_0^{+\infty}[z^T(t)z(t)-\gamma^2 w^T(t)w(t)]dt$$
$$\le\int_0^{+\infty}[\dot{V}(x(t),w(t),t)+z^T(t)z(t)-\gamma_i^2 w^T(t)w(t)]dt-V(x(+\infty),w(+\infty),+\infty)$$
$$=\sum_{i=1}^{M}\int_{t_{i-1}}^{t_i}[\dot{V}(x(t),w(t),t)+z^T(t)z(t)-\gamma_i^2 w^T(t)w(t)]dt-V(x(+\infty),w(+\infty),+\infty)<0$$

Therefore

$$\int_0^{\infty}z^T(t)z(t)dt\le\gamma^2\int_0^{+\infty}w^T(t)w(t)dt$$

This completes the proof.

Remark 3. Theorem 4 presents a sufficient condition for robust stabilization with H_∞ performance γ. Since there are uncertainties in inequality (43), it can not be solved directly.

Lemma 5：For system (39), and given positive scalar γ_i, if there are matrices $Y_i, Y_{1i}, ...Y_{Ni}$, positive definite matrices $X_i, R_{1i}, ..., R_{Ni}$ with proper dimensions, and scalar $\alpha > 0$, such that the following inequality holds

$$\begin{bmatrix} S & A_{d1}X_i + BY_{1i} & ... & A_{dN}X_i + BY_{Ni} & B_1 & \varphi & X_iE_1^T + Y_i^TE_3^T \\ * & -R_{1i} & 0 & 0 & 0 & X_iC_{d1}^T + Y_{1i}^TD^T & X_iE_{21}^T + Y_{1i}^TE_3^T \\ * & * & ... & 0 & 0 & ... & ... \\ * & * & * & -R_{Ni} & 0 & X_iC_{dN}^T + Y_{Ni}^TD^T & X_iE_{2N}^T + Y_{Ni}^TE_3^T \\ * & * & * & * & -\gamma_i^2 I & B_2^T & 0 \\ * & * & * & * & * & -I + \alpha H_2H_2^T & 0 \\ * & * & * & * & * & * & -\alpha I \end{bmatrix} < 0 \tag{48}$$

where $S = X_iA^T + AX_i + Y_i^TB^T + BY_i + \sum_{j=1}^{N} R_{ji} + \alpha H_1H_1^T$, $\varphi = X_iC^T + Y_i^TD^T + \alpha H_1H_2^T$,the system (39) is robust stabilizable with H_∞ performance $\gamma = \max(\gamma_i)$. The robust H_∞ control is given by：

$$u(t) = K_ix(t) + \sum_{j=1}^{N} K_{ji}x(t-\tau_{ji}), K_i = Y_iX_i^{-1}, K_{ji} = Y_{ji}X_i^{-1}, i \in M, j \in N$$

The switching law is $\sigma(t) = \arg\min_{i \in M}\{x^T(t)X_i^{-1}x(t)\}$。

Proof. The proof is similar to Theorem 3, and is omitted.

Example 3 The linear uncertain switched system (39) with multiple time delays is given below.

$$\begin{aligned} A^1 &= \begin{bmatrix} -1 & 1 \\ 0 & 1 \end{bmatrix}, A_1^1 = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}, A_2^1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, B^1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, B_1^1 = \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix} \\ B_2^1 &= \begin{bmatrix} 0.2 \\ 0.1 \end{bmatrix}, C^1 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, C_1^1 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0 \end{bmatrix}, C_2^1 = \begin{bmatrix} 0 & 0 \\ 0 & 0.2 \end{bmatrix}, D^1 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.4 \end{bmatrix} \\ H_1^1 &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, H_2^1 = \begin{bmatrix} 0.5 & 0.4 \\ 0 & 0 \end{bmatrix}, E_1^1 = \begin{bmatrix} 0.5 & 0.2 \\ 0 & 0 \end{bmatrix}, E_{21}^1 = \begin{bmatrix} 0.5 & 0.4 \\ 0 & 0 \end{bmatrix}, E_{22}^1 = \begin{bmatrix} 0.2 & 0.5 \\ 0 & 0 \end{bmatrix} \\ E_3^1 &= \begin{bmatrix} 0.4 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{49}$$

and $\gamma^1 = 3, w^1(t) = \cos(t), F^1(t) = \sin(t), \alpha^1 = 0.5, \tau_1^1 = \tau_2^1 = 0.02$

$$\begin{aligned} A^2 &= \begin{bmatrix} -1 & 0 \\ 1 & 1 \end{bmatrix}, A_1^2 = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}, A_2^2 = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}, B^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, B_1^2 = \begin{bmatrix} 0.2 \\ 0.1 \end{bmatrix} \\ B_2^2 &= \begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix}, C^2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, C_1^2 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0 \end{bmatrix}, C_2^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0.2 \end{bmatrix}, D^2 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.4 \end{bmatrix} \\ H_1^2 &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, H_2^2 = \begin{bmatrix} 0.3 & 0.4 \\ 0 & 0 \end{bmatrix}, E_1^2 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.2 \end{bmatrix}, E_{21}^2 = \begin{bmatrix} 0.2 & 0.4 \\ 0 & 0 \end{bmatrix}, E_{22}^2 = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.2 \end{bmatrix} \\ E_3^2 &= \begin{bmatrix} 0.5 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

and $\gamma^2 = 1, w^2(t) = \cos(t), F^2(t) = \sin(t), \alpha^2 = 0.5, \tau_1^2 = \tau_2^2 = 0.02$

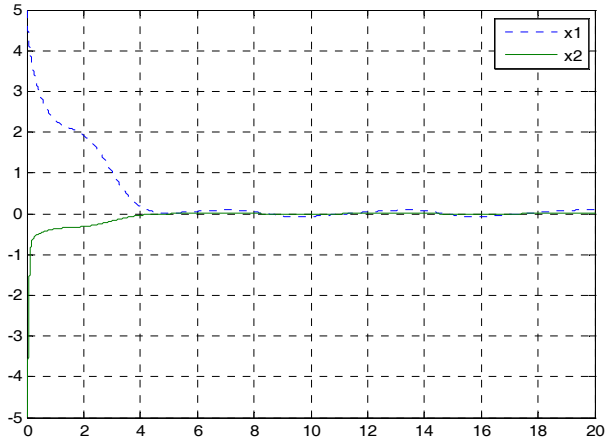


Fig. 3. State response of Example 3

By Lemma 5, the state feedback gain for system stabilization with H_∞ performance are

$$P^1 = X_1^{-} = \begin{bmatrix} 0.0352 & 0.2345 \\ 0.2345 & 1.5610 \end{bmatrix}, K^1 = \begin{bmatrix} -0.0171 & -0.1135 \\ -0.3927 & -2.6146 \end{bmatrix}, K_1^1 = \begin{bmatrix} 0.0559 & 0.3723 \\ -1.0425 & -6.9413 \end{bmatrix}, K_2^1 = \begin{bmatrix} 0.0559 & 0.3723 \\ -1.0425 & -6.9413 \end{bmatrix}$$

$$P^2 = X_2^{-} = \begin{bmatrix} 0.0025 & -0.0842 \\ -0.0842 & 2.6723 \end{bmatrix}, K^2 = \begin{bmatrix} -0.0075 & 0.2345 \\ 0.1130 & -3.6644 \end{bmatrix}, K_1^2 = \begin{bmatrix} -0.0714 & 2.3169 \\ 0.2707 & -8.7784 \end{bmatrix}, K_2^2 = \begin{bmatrix} -0.0721 & 2.3395 \\ 0.2729 & -8.8480 \end{bmatrix}$$

The switching law is:

$$\sigma(t) = \mathrm{i} = \begin{cases} 1, x^T(P^1 - P^2)x \le 0 \\ 2, x^T(P^1 - P^2)x > 0 \end{cases}$$

The state responses are shown in Figure 3.

x1 and x2 are system states, the initial condition is [x1,x2]=[5,-5]。The result shows the system is stable under the switching law when it is switched among the closed-loop subsystems.

6. Conclusion

This chapter studies the robust H_∞ control for linear switched systems with time delay. After introducing robust H_∞ stability and stabilization of linear switched systems, we firstly analyzed robust H_∞ control for general linear switched systems with time delay. Based on the multi-Lyapunov-Function method, a sufficient condition is derived in terms of LMI. The robust H_∞ control and the switching law design are also given.

By involving uncertainties and time-varying delay, the robust $H\infty$ control for uncertain linear switched systems with time varying delay is studied. Through the multi-Lyapunov-Function approach, a sufficient condition is given in LMI formulation. The robust switched $H\infty$ control and the the switching law design are presented as well.

The state feedback robust $H\infty$ control with memory is also studied for uncertain linear switched systems with multiple time delays. A sufficient condition is given as well as the robust switched $H\infty$ control and the switching law.

Illustrative examples are given to show the effectiveness of the proposed methods.

7. References

Wu, L.; & Meng, W.X. (2009). Weighted $H\infty$ model reduction for linear switched systems with time-varying delay, *Automatic*, Vol. 45, No. 1, (Jan. 2009), 186-193, ISSN: 00051098

Zhang, X.; & Liu, Y. (2008). Delay-dependent robust $H\infty$ control of a class of uncertain switched systems with time delay, *Proceedings of the 7th World Congress on Intelligent Control and Automation*, pp. 8873-8877, ISBN-13: 9781424421145, Chongqing, China, June 2008, IEEE, Piscataway

Xie, D.; Wang, L.; Hao, F.; & Xie, G. (2004). LMI approach to L_2-gain analysis and control synthesis of uncertain switched systems. *IEEE Proc.-Control Theory Appl.*, Vol. 151, No. 1, (Jan. 2004), 21-28, 13502379

Fu, Z.; Fei, S.; Long, F. & Cong, S. (2007). Robust $H\infty$ dynamic output feedback stabilization for a class of uncertain switched systems. *Journal of Southeast University: Natural Science*, Vol. 37, No. 1, (Jan. 2007), 1304-1310, 10010505

Song, Z.; Nie, H.; & Zhao, J. (2007). Robust $H\infty$ control of discrete-Time Switched Systems with time-varying delay**.** *Journal of Northeastern University: Natural Science,* Vol. 28, No. 4, (Apr. 2007), 469-472, 10053026

Song, Z. & Zhao, J. (2006) Robust control of uncertain discrete-time switched systems with time-delay. *Acta Automatic Sinica*, Vol. 32, No. 5, (May 2006), 760-766, 02544156

Ma, Y. ; Yan W. ; Liu D. & Zheng Y. (2006) $H\infty$ control of discrete switched systems with time delay via state feedback with memory[C]. *Proceedings of the 6th World Congress on Intelligent Control and Automation*, pp. 1328-1332, ISBN-13: 9781424403325, Dalian, China, 2006, IEEE, Piscataway

5

Active Suspension in Integrated Vehicle Control

Péter Gáspár, Zoltán Szabó and József Bokor
Computer and Automation Research Institute
Hungary

1. Introduction

These days road vehicles contain several individual active control mechanisms that solve a large number of control tasks. These components are often highly nonlinear, which are modelled as hybrid systems. An example is the semiactive/active suspension system, which can be modelled as a nonlinear dynamics augmented with an actuator that has a bimodal dynamics, i.e. a closed loop switching system with two modes. Moreover, in traditional control systems the vehicle functions to be controlled are designed and implemented separately. Although in the design of the individual control components only a subset of the full vehicle dynamics is considered these components influence the entire vehicle. Thus in the operation of these autonomous control systems interactions and conflicts may occur that might overwrite the intentions of the designers concerning the individual performance requirements. The aim of the integrated control methodologies is to combine and supervise all controllable subsystems affecting vehicle dynamic responses in order to ensure the management of resources. The solution might be the integration of the control logic of subsystems.
Active suspensions are used to provide good handling characteristics and improve ride comfort while harmful vibrations caused by road irregularities and on-board excitation sources act upon the vehicle. The performance of suspension systems is assessed quantitatively in terms of several parameters: passenger comfort, suspension deflection, tire load variation and energy consumption, see Gillespie (1992); Sharp & Crolla (1987). In order to improve passenger comfort it is important to keep the effects of the road disturbance on the heave acceleration small. Structural features of the vehicle place a hard limit on the amount of suspension deflection available for reducing the acceleration of the vehicle body. Hence it is also important to keep the effect of the disturbance on the suspension deflection sufficiently small. In order to reduce the dynamic tire load deflection, the effects of the disturbance on tire deflection should also be kept small. The control force limitation is incorporated into the design procedure in order to avoid large control forces.
Applying a braking force decelerates the vehicle. Additionally, the role of the active brake is to apply unilateral braking since it reduces the lateral tire forces directly Chen & Peng (2001); Palkovics et al. (1999). This feature provides a redundancy in affecting the lateral dynamics. However, using the active brake might have unwanted side effects as the modification of the yaw dynamics and of the longitudinal direction of the vehicle. Therefore the use of the active brake is preferred only in emergencies.
In this paper a control structure that integrates active suspensions and an active brake is proposed to improve the safety of vehicles. On the global level the active suspension system is primarily designed to improve passenger comfort and road holding. However, it is able to

generate a stabilizing moment to balance an overturning moment during vehicle maneuvers in order to reduce the rollover risk. The controlled braking system is activated only when the vehicle comes close to rolling over. During abrupt brakings pitch dynamics increases significantly. The active suspension is also able to generate a moment and improve the pitch stability of the vehicle.
Rollover prevention is a safety feature. The combination of the active suspension and active brake handles the emergency situation provided that the active suspension component is fully functional. In order to enhance safety the reconfigurable control is extended with a fault-tolerant property in order to guarantee performances even if a hydraulic actuator fault occurs in the active suspension system. The solution of the fault-tolerant operation requires the reconfigurability of the active brake.
The control design of switched systems is involved at two levels: at the suspension actuator level a tracking controller is designed for a bimodal system together with a fault detection filter.
This paper presents the application of the Linear Parameter Varying (LPV) method for the design of integrated vehicle control systems, in which several active components are used in co-operation. In the control design besides performance specifications and uncertainties, the fault information can be taken into consideration. By monitoring suitable scheduling parameters in the LPV control, the reconfiguration of the control systems can be achieved, conflict between performance demands can be avoided and faults (loss in effectiveness) can be handled. This level provides the reference signal for the low-level actuator design and it also constitutes the supervisor controller for the reconfiguration. By using the LPV method the designed controller guarantees the desired stability and performance demands of the closed–loop system. The operation of the control systems is demonstrated through various simulation vehicle maneuvers.
The structure of the paper is the following: the global chassis model containing both the vertical and later dynamics is presented in Section 2. It is followed by a detailed formulation of the control problem in Section 3. The actuator dynamics is considered in Section 4. where a tracking control is designed. Possible faults of the suspension actuator are detected by using the FDI filter of Section 5. The proposed method is demonstrated through a series of simulation examples in Section 6. Finally some conclusion remarks are formulated in Section 7.

2. Global chassis model

The class of finite dimensional linear systems, whose state space entries depend continuously on a time varying parameter vector, $\rho(t)$, is called LPV. The trajectory of the vector-valued signal, $\rho(t)$ is assumed not to be known in advance, although its value is accessible (measured) in real time and is constrained a priori to lie in a specified bounded set. The idea behind using LPV systems is to take advantage of the casual knowledge of the dynamics of the system, see Becker & Packard (1994); Leith & Leithead (2000); Rough & Shamma (2000); Wu (2001). The formal definition of an LPV system is given below:
For a compact subset $\mathcal{P} \subset \mathcal{R}^S$, the parameter variation set $\mathcal{F}_\mathcal{P}$ denotes the set of all piecewise continuous functions mapping $\mathcal{R}$ (time) into $\mathcal{P}$ with a finite number of discontinuities in any interval. The compact set $\mathcal{P} \subset \mathcal{R}^S$, along with continuous functions $A : \mathcal{R}^S \to \mathcal{R}^{n\times n}$ and $B : \mathcal{R}^S \to \mathcal{R}^{n\times n_u}$ represents an n^{th} order LPV system $G(\rho)$ whose dynamics evolve as

$$\dot{x} = A(\rho)x + B(\rho)u, \tag{1}$$

where $\rho \in \mathcal{F}_{\mathcal{P}}$. One characteristics of the LPV system is that it must be linear in the pair formed by the state vector x, and the control input vector u. The matrices A and B are generally nonlinear functions of the scheduling vector ρ.
Due to the complexity of a vehicle model two models are formalized. One model is suitable for designing the suspension system and it also takes the vertical dynamics into consideration. The other model is used to design lateral dynamics and the brake system. Connection between the two models is achieved by the application of the integrated control system.
In order to describe the vertical dynamics a full-car model, which is shown in Figure 1, is used that comprises five parts: the sprung mass and four unsprung masses at the left and right hand side at the front and rear. All the suspensions are modelled as an ensemble of a spring, a damper and an actuator to generate a pushing force between the body and the axle. The suspension stiffness and the tire stiffness are denoted by k_s and $k_t$while front and rear suspension dampers are denoted by b_s,$respectively$. Let the front and rear displacement of the sprung mass on the left and right side be denoted by x_{1fl}, x_{1rl} and x_{1fr}, x_{1rr}. Let the front and rear displacement of the unsprung mass on the left and right side be denoted by x_{2fl}, x_{2rl}, x_{2fr}, and x_{2rr}. In the full-car model, the disturbances, w_{fl}, w_{rl}, w_{fr}, w_{rr} are caused by road irregularities. The input signals, f_{fl}, f_{rl}, f_{fr}, f_{rr} are generated by the actuators. The system

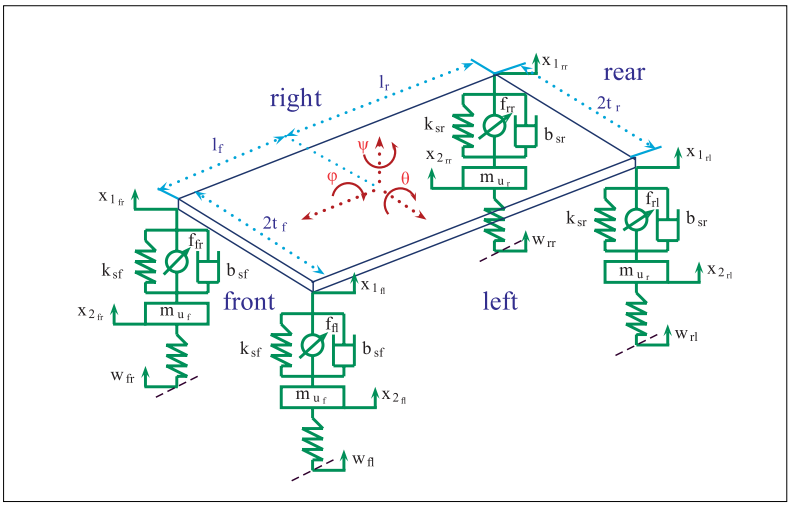


Fig. 1. Vertical model of the vehicle

equations correspond to a seven degrees-of-freedom full-car vehicle model. The sprung mass is assumed to be a rigid body and has freedoms of motion in the vertical, pitch and roll directions. The x_1 is the vertical displacement at the center of gravity, θ is the pitch angle and ϕ is the roll angle of the sprung mass, respectively. Each unsprung mass has freedom of motion in the vertical direction, x_{2fl}, x_{2rl}, x_{2fr}, x_{2rr}. The vehicle dynamical model, i.e. the heave motion,

the pitch motion, the roll motion, the front and rear tires, is as follows:

$$\begin{aligned} m_s\ddot{x}_1 &= F_{kfl} + F_{kfr} + F_{krl} + F_{krr} + F_{bfl} + F_{bfr} + F_{brl} + F_{brr} \\ &\quad - f_{fl} - f_{fr} - f_{rl} - f_{rr}, \end{aligned} \tag{2}$$

$$\begin{aligned} I_\theta\ddot{\theta} &= l_f F_{kfl} + l_f F_{kfr} - l_r F_{krl} - l_r F_{krr} + l_f F_{bfl} + l_f F_{bfr} - l_r F_{brl} + l_r F_{brr} \\ &\quad - l_f f_{fl} - l_f f_{fr} + l_r f_{rl} + l_r f_{rr}, \end{aligned} \tag{3}$$

$$\begin{aligned} I_\phi\ddot{\phi} &= t_f F_{kfl} - t_f F_{kfr} + t_r F_{krl} - t_r F_{krr} + t_f F_{bfl} - t_f F_{bfr} + t_r F_{brl} - t_r F_{brr} \\ &\quad - t_f f_{fl} + t_f f_{fr} - t_r f_{rl} + t_r f_{rr}, \end{aligned} \tag{4}$$

$$m_{uf}\ddot{x}_{2_{fl}} = -F_{kfl} - F_{tfl} - F_{bfl} + f_{fl}, \tag{5}$$

$$m_{uf}\ddot{x}_{2_{fr}} = -F_{kfr} - F_{tfr} - F_{bfr} + f_{fr}, \tag{6}$$

$$m_{ur}\ddot{x}_{2_{rl}} = -F_{krl} - F_{trl} - F_{brl} + f_{rl}, \tag{7}$$

$$m_{ur}\ddot{x}_{2_{rr}} = -F_{krr} - F_{trr} - F_{brr} + f_{rr}, \tag{8}$$

where the following linear approximations are applied:

$$\begin{aligned} x_{1_{fl}} &= x_1 + l_f\theta + t_f\phi, \qquad x_{1_{fr}} = x_1 + l_f\theta - t_f\phi, \\ x_{1_{rl}} &= x_1 - l_r\theta + t_r\phi, \qquad x_{1_{rr}} = x_1 - l_r\theta - t_r\phi. \end{aligned}$$

The suspension damping force and the suspension spring force, respectively, are as follows:

$$F_{bij} = b_s^l(\dot{x}_{2ij} - \dot{x}_{1ij}) - b_s^{sym}|\dot{x}_{2ij} - \dot{x}_{1ij}| + b_s^{nl}\sqrt{|\dot{x}_{2ij} - \dot{x}_{1ij}|}\,\mathrm{sgn}(\dot{x}_{2ij} - \dot{x}_{1ij}), \tag{9}$$

$$F_{kij} = k_s^l(x_{2ij} - x_{1ij}) + k_s^{nl}(x_{2ij} - x_{1ij})^3, \tag{10}$$

and f_{ij} are the forces of the actuator, where $ij \in \{fl, fr, rl, rr\}$. Here, parts of the nonlinear suspension damper b_s are b_s^l, b_s^{nl} and b_s^{sym}. The b_s^l coefficient affects the damping force linearly while b_s^{nl} has a nonlinear impact on the damping characteristics. b_s^{sym} describes the asymmetric behavior of the characteristics. Parts of the nonlinear suspension stiffness k_s are a linear coefficient k_s^l and a nonlinear one, k_s^{nl}. The tire force is approximated by a linear model:

$$F_{tij} = k_t(x_{2ij} - w_{ij}). \tag{11}$$

The state vector x is selected as follows:

$$x_s = \begin{bmatrix} q & x_u & \dot{q} & \dot{x}_u \end{bmatrix}^T \tag{12}$$

with $q = \begin{bmatrix} x_1 & \theta & \phi \end{bmatrix}^T$ and $x_u = \begin{bmatrix} x_{2fl} & x_{2fr} & x_{2rl} & x_{2rr} \end{bmatrix}^T$. The state space representation of the LPV model is as follows:

$$\dot{x}_s = A_s(\rho_s)x_s + B_{1s_v}(\rho_s)d_s + B_{2s_v}(\rho_s)u_s, \tag{13}$$

where

$$u_s = \begin{bmatrix} f_{fl} & f_{fr} & f_{rl} & f_{rr} \end{bmatrix}^T. \tag{14}$$

The disturbance is $d_s = \begin{bmatrix} w_{fl} & w_{rl} & w_{fr} & w_{rr} \end{bmatrix}$. Variables concerning the front and rear displacement between the sprung mass and the unsprung mass on the left and right side and their velocities are selected as scheduling variables:

$$\rho_s = \begin{bmatrix} \rho_{bij} & \rho_{kij} \end{bmatrix}^T, \qquad ij \in (fl, fr, rl, rr) \tag{15}$$

where

$$\rho_{bij} = \dot{x}_{2ij} - \dot{x}_{1ij}, \tag{16}$$

$$\rho_{kij} = x_{2ij} - x_{1ij} \tag{17}$$

The scheduling variables ρ_{bij} depend on the relative velocity, while the scheduling variables ρ_{kij} depend on the relative displacement. In practice, the relative displacement is a measured signal. The relative velocity is then determined by numerical differentiation from the measured relative displacement.

Figure 2 illustrates the combined yaw-roll dynamics of the vehicle modelled by a three-body system, in which m_s is the sprung mass, $m_{u,f}$ is the unsprung mass at the front including the front wheels and axle, and $m_{u,r}$ is the unsprung mass at the rear with the rear wheels and axle. β denotes the side slip angle of the sprung mass, ψ is the heading angle, ϕ is the roll angle, $\dot{\psi}$ denotes the yaw rate and θ the pitch angle. The roll angle of the unsprung mass at the front and at the rear axle are denoted by $\phi_{t,f}$ and $\phi_{t,r}$, respectively. a_y denotes the lateral acceleration and z_s is the heave displacement while v stands for the forward velocity.

In the vehicle modelling the motion differential equations of the yaw-roll dynamics of the single unit vehicle, i.e. the lateral dynamics, the yaw moment, the roll moment of the sprung mass, the roll moment of the front and the rear unsprung masses, are formalized.

$$mv(\dot{\beta} + \dot{\psi}) - m_s h\ddot{\phi} = F_{y,f} + F_{y,r}, \tag{18}$$

$$-I_{xz}\ddot{\phi} + I_{zz}\ddot{\psi} = F_{y,f} l_f - F_{y,r} l_r + l_w \Delta F_b, \tag{19}$$

$$\begin{aligned} (I_{xx} + m_s h^2)\ddot{\phi} - I_{xz}\ddot{\psi} = {} & m_s g h \phi + m_s v h (\dot{\beta} + \dot{\psi}) \\ & - k_f(\phi - \phi_{t,f}) - b_f(\dot{\phi} - \dot{\phi}_{t,f}) + u_f \\ & - k_r(\phi - \phi_{t,r}) - b_r(\dot{\phi} - \dot{\phi}_{t,r}) + u_r, \end{aligned} \tag{20}$$

$$\begin{aligned} -h_r F_{y,f} = {} & m_{u,f} v (h_r - h_{u,f})(\dot{\beta} + \dot{\psi}) + m_{u,f} g h_{u,f} \phi_{t,f} - k_{t,f} \phi_{t,f} \\ & + k_f(\phi - \phi_{t,f}) + b_f(\dot{\phi} - \dot{\phi}_{t,f}) + u_f, \end{aligned} \tag{21}$$

$$\begin{aligned} -h_r F_{y,r} = {} & m_{u,r} v (h_r - h_{u,r})(\dot{\beta} + \dot{\psi}) - m_{u,r} g h_{u,r} \phi_{t,r} - k_{t,r} \phi_{t,r} \\ & + k_r(\phi - \phi_{t,r}) + b_r(\dot{\phi} - \dot{\phi}_{t,r}) + u_r. \end{aligned} \tag{22}$$

The detailed derivation of the equations of the yaw-roll dynamics of the single unit vehicle can be found in Sampson & Cebon (2003).

The lateral tire forces $F_{y,i}$ in the direction of velocity at the wheel ground contact points are approximated proportionally to the tire side slip angle α_i:

$$F_{y,f} = \mu C_f \alpha_f, \qquad F_{y,r} = \mu C_r \alpha_r.$$

The C_i is the tire side slip constant and α_i is the tire side slip angle associated with the front and rear axles. The chassis and the wheels have identical velocities at the wheel ground contact points. The velocity equations for the front and rear wheels in the lateral and in the

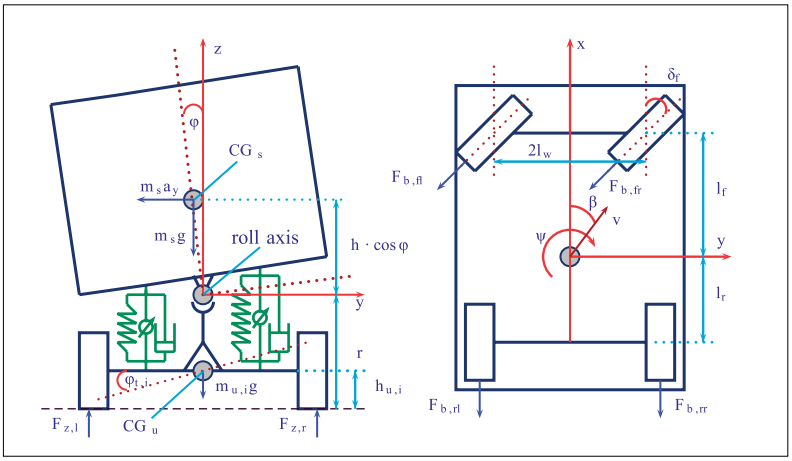


Fig. 2. Yaw and roll model of the vehicle

longitudinal directions are as follows:

$$v_{w,f}\sin(\delta_f-\alpha_f)=l_f\cdot\dot{\psi}+v\sin\beta,\qquad v_{w,f}\cos(\delta_f-\alpha_f)=v\cos\beta,$$
$$v_{w,r}\sin\alpha_r=l_r\cdot\dot{\psi}-v\sin\beta,\qquad v_{w,r}\cos\alpha_r=v\cos\beta.$$

In stable driving conditions, the tire side slip angle α_i is normally not larger than five degrees and the above equation can be simplified by substituting $\sin x\approx x$ and $\cos x\approx 1$. The classic equations for the tire side slip angles are then given as

$$\alpha_f=-\beta+\delta_f-\frac{l_f\cdot\dot{\psi}}{v},\qquad \alpha_r=-\beta+\frac{l_r\cdot\dot{\psi}}{v}.$$

The equations (18)-(22) can be expressed in the state space representation form. The system states are the side slip angle of the sprung mass β, the yaw rate $\dot{\psi}$, the roll angle ϕ, the roll rate $\dot{\phi}$, the roll angle of the unsprung mass at the front axle $\phi_{t,f}$ and at the rear axle $\phi_{t,r}$. Let the state vector be the following:

$$x_r=\begin{bmatrix}\beta & \dot{\psi} & \phi & \dot{\phi} & \phi_{t,f} & \phi_{t,r}\end{bmatrix}^T. \tag{23}$$

Using the state vector, the differential algebraic model defined by Equations (18)-(22) is transformed into a state space representation form:

$$\dot{x}_r=A_r(\rho_r)x_r+B_{1r_v}(\rho_r)d_r+B_{2r_v}(\rho_r)u_r. \tag{24}$$

The disturbance is the front wheel steering angle: $d_r=\delta_f$, while the control inputs are set to be:

$$u_r=\begin{bmatrix}\Delta F_b & u_{af} & u_{ar}\end{bmatrix}^T. \tag{25}$$

In this approach of the rollover problem the active suspensions generate two stabilizing roll moments at the front and the rear, which can be considered as the effects of the suspension forces

$$u_{af} = (f_{fl} - f_{fr})\ell_w, \qquad u_{ar} = (f_{rl} - f_{rr})\ell_w. \tag{26}$$

The roll moments required are distinguished equally at the suspension components:

$$f_{fl} = \frac{u_{af}}{2l_w}, \qquad f_{fr} = -\frac{u_{af}}{2l_w}, \tag{27}$$

$$f_{rl} = \frac{u_{ar}}{2l_w}, \qquad f_{rr} = -\frac{u_{ar}}{2l_w} \tag{28}$$

The third control input is the difference in brake forces between the left and right-hand sides of the vehicle:

$$\Delta F_b = (F_{brl} + d_2 F_{bfl}) - (F_{brr} + d_1 F_{bfr}), \tag{29}$$

where d_1 and d_2 are distances, which depend on the steering angle. In the implementation of the controller means that the control action be distributed at the front and the rear wheels at either of the two sides.
The differential equations depend on the forward velocity v and the adhesion coefficient μ of the vehicle nonlinearly. It is assumed that the forward velocity and the adhesion coefficient are available, i.e. these parameters are estimated on-line by using the on-board sensors. A grey-box identification method based on an observer design was proposed in Gáspár et al. (2006). The scheduling vector ρ_r is selected with four scheduling variables

$$\rho_r = \begin{bmatrix} \rho_1 & \rho_2 & \rho_3 & \rho_4 \end{bmatrix} \tag{30}$$

with $\rho_1 = \mu$,$\rho_2 = \frac{\mu}{v}$,$\rho_3 = \frac{\mu}{v^2}$ and $\rho_4 = \frac{1}{v}$.

3. Supervisory global control

The control design for suspension system and rollover prevention is performed on a full-car vehicle model. When a fault occurs in the active suspension system, its role is assumed by the active brake.The orchestration of the two independent subsystems, i.e. the suspension subsystem and active brake, respectively, should be solved by a dedicated mechanism in order to guarantee a desired level of the required performance.
The detection of an imminent rollover is based on the monitoring of the lateral load transfers for both axles. The lateral load transfer can be given:

$$\Delta F_{z,i} = \frac{k_{t,i}\phi_{t,i}}{l_w}, \tag{31}$$

where i denotes the front and rear axles. They can be normalized in such a way that the load transfer is divided by the total axle load:

$$R_i = \frac{\Delta F_{z,i}}{F_{z,i}}. \tag{32}$$

The normalized load transfer R_i value corresponds to the largest possible load transfer. If the R_i takes on the value ± 1 then the inner wheels in the bend lift off. The limit cornering

condition occurs when the load on the inside wheels has dropped to zero and all the load has been transferred onto the outside wheels. Let

$$R = max\{R_f, R_r\}. \tag{33}$$

The roll angles of the unsprung masses have an important role in the monitoring of rollovers, since the calculation of the normalized load transfers is based on these signals.
In the control design both the rollover and the suspension problems are taken into consideration. In this combined structure a new weighting strategy is proposed in order to meet several performance demands, such as enhancing passenger comfort, increasing rollover stability and road holding, guaranteing suspension working space and reducing energy consumption. In the rollover problem the performance outputs for control design are the lateral acceleration, the lateral load transfers at the front and the rear, and the control inputs:

$$z_r = \begin{bmatrix} a_y & \Delta F_{z,f} & \Delta F_{z,r} & u_r \end{bmatrix}^T. \tag{34}$$

In the suspension problem the performance outputs

$$z_s = \begin{bmatrix} a_z & z_{sf} & z_{sr} & u_s \end{bmatrix}^T \tag{35}$$

for control design are the passenger comfort (i.e. heave acceleration), the suspension deflections and the control inputs. The measured outputs are the lateral acceleration of the sprung mass, the derivative of the roll angle and the suspension deflections at the suspension components: $y_r = \begin{bmatrix} a_y & \dot{\phi} \end{bmatrix}^T$ and $y_s = \begin{bmatrix} z_{sf} & z_{sr} \end{bmatrix}^T$.
In order to achieve the desired reconfiguration of the redundant subsystems a straight solution would be to apply merely a switching strategy that would change between the two subsystems in emergency, i.e. when an imminent rollover occur. By applying a switching strategy based on a suitable threshold imposed for the value of the normalized load transfer R (switching surface) would be a reliable solution for the design of the individual controllers. However, it would generate the problem of the transients during the switching instances, i.e. the required performance level should be imposed by special techniques in that case.

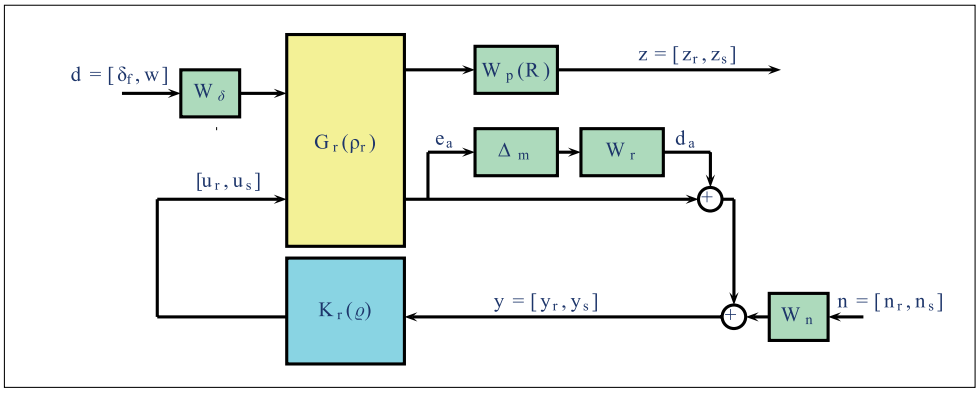


Fig. 3. The closed-loop interconnection structure for the control design

Instead of considering a switching surface the reconfiguration of the control structure is solved by an LPV strategy, which is presented in this section. The values of the the normalized load

transfer R are used as a scheduling variable for a LPV design where a guaranteed performance level during the reconfiguration is achieved through the design process – through a common Lyapunov function. This is possible by using a suitable weighting strategy where the critical parameter R schedules the performance weight functions.
The closed-loop interconnection structure for rollover prevention and for suspension design are shown in Figure 3. The purpose of the weighting functions is to keep the lateral acceleration, the lateral load transfers, the heave acceleration, the suspension deflection and the control inputs small over the desired frequency range. The weighting functions chosen for performance outputs can be considered as penalty functions: they are selected large in a frequency range where small signals are desired, and small where larger performance outputs can be tolerated.
The weighting function for the lateral acceleration, for the heave acceleration and for the suspension deflection are selected in the following way:

$$W_{p,ay} = \phi_{ay}(R)\frac{A_1(\frac{s}{T_a}+1)}{(\frac{s}{T_b}+1)}, \tag{36}$$

$$W_{p,az} = \frac{A_2(\frac{s}{T_c}+1)}{(\frac{s}{T_d}+1)}, \tag{37}$$

$$W_{p,zs} = \frac{A_3(\frac{s}{T_e}+1)}{(\frac{s}{T_f}+1)}, \tag{38}$$

respectively, with time constants T_i and proportional coefficients A_i. The weighting functions $W_{p,us}$ and $W_{p,ur}$ for the control inputs guarantee the limitation of the control forces.
For safety the weighting function $W_{p,ay}$ for the lateral acceleration plays the most important role. The parameter-dependent gain $\phi_{ay}(R)$ in the weighting function $W_{p,ay}$ is selected as a function of parameter R:

$$\phi_{ay}(R) = \begin{cases} 0 & \text{if} \quad |R| < R_a \\ \frac{(|R|-R_a)}{(R_b-R_a)} & \text{if} \quad R_a \leq |R| \leq R_b \\ 1 & \text{if} \quad |R| > R_b \end{cases} \tag{39}$$

where R_a, R_b are pre-defined and constants in fault-free case.
The gain ϕ_{ay} is increased in order to minimize the lateral acceleration and prevent the rollover of the vehicle. As the gain ϕ_{ay} increases the lateral acceleration decreases, since the active brake influences the lateral acceleration directly. R_a defines the critical status when the vehicle is in an emergency. Parameter R_b shows how fast the control should focus on minimizing the lateral acceleration. In the lower range of R the gain ϕ_{ay} must be small, and in the upper range of R the gains must be large. Consequently, the weighting functions must be selected in such a way that they minimize the lateral load transfers in emergency situations. However in normal cruising situations the control do not focus on the lateral load transfers since the weight is small.
The suspension system reduces the rollover risk when $|R| \geq R_a$. The suspension forces are modified by the fictitious forces coming from the stabilizing moments, see equation In an emergency, i.e. when $|\rho_R| \geq R_s$, the suspension system must reduce the rollover risk and guaranteeing passenger comfort (and pitch angle) is no longer a priority. The suspension forces are modified by the fictitious forces coming from the stabilizing moments, see equation

(27). The forces at the front and the rear in both sides are the following:

$$f_{fl,new} = f_{fl} + \frac{u_{rf}}{2l_w}, \quad f_{fr,new} = f_{fr} - \frac{u_{rf}}{2l_w}, \tag{40}$$

$$f_{rl,new} = f_{rl} + \frac{u_{rr}}{2l_w}, \quad f_{rr,new} = f_{rr} - \frac{u_{rr}}{2l_w}. \tag{41}$$

In the event of a fault the range of the operation of the brake system must be extended and the wheels are decelerated gradually rather than rapidly if the normalized load transfer has reached its critical value. A small value of R_a corresponds to activating the brake system early and gradually, whereas a large value of R_a corresponds to activating the brake system rapidly. Thus, the design parameter R_a is chosen to be scheduled on fault information ρ_f.

$$R_{a,new} = R_a - \frac{\rho_f}{\alpha} \tag{42}$$

where ρ_f is the normalized value of the fault information and α is a constant factor.
Introducing the health information about the suspension subsystem (through the fault signal ρ_f) in the control design has the benefit in reducing the brake actuation during operational time. This is achieved by a varying threshold of the activation level, which is R_a for the fault–free case and it decreases only as an exception when a loss of effectiveness/fault occur in contrast to the conservative setting $R_a - \frac{1}{\alpha}$ when no information about the suspension is supposed to be available.
The uncertainties of the nominal model are represented by the weighting function W_r in such a way that in the low frequency domain the uncertainties are about 10% and in the upper frequency domain they are up to 100%. The input scaling weights W_δ and W_w normalize the disturbances to the maximum expected command. $W_{n,ay}$, $W_{n,\dot{\phi}}$, $W_{n,\dot{\psi}}$ and $W_{n,sij}$ take into account the sensor noises in the control design.
In order to describe the control objective, the parameter dependent augmented plant $P(\varrho)$ must be built up using the closed-loop interconnection structure. The augmented plant $P(\varrho)$ includes the parameter dependent vehicle dynamics and the weighting functions.

$$\left[\begin{array}{c} \tilde{z} \\ \hline y \end{array}\right] = \left[\begin{array}{c|c} P_{11}(\varrho) & P_{12}(\varrho) \\ \hline P_{21}(\varrho) & P_{22}(\varrho) \end{array}\right] \left[\begin{array}{c} w \\ \hline u \end{array}\right], \tag{43}$$

where $w = \begin{bmatrix} d & n & d_m \end{bmatrix}$ and $\tilde{z} = \begin{bmatrix} z & e_m \end{bmatrix}$. The signals d_m, e_m are the output of the uncertainty block Δ_m and its input, respectively. ϱ contains both the scheduling variables from the vehicle modeling and the monitoring variables. The closed-loop system $M(\varrho)$ is given by a lower linear fractional transformation (LFT) structure:

$$M(\varrho) = \mathcal{F}_\ell(P(\varrho), K(\varrho)), \tag{44}$$

where $K(\varrho)$ depends on the scheduling parameter ϱ. The purpose of the control design is to robustly minimize the induced $\mathcal{L}_2$ norm of a LPV system $M(\varrho)$ with zero initial conditions, which is given by

$$\inf_K \sup_{\varrho \in \mathcal{F}_\mathcal{P}} \sup_{\|w\|_2 \neq 0, w \in \mathcal{L}_2} \frac{\|\tilde{z}\|_2}{\|w\|_2} \tag{45}$$

The solution of an LPV problem is based on the set of infinite dimensional LMIs being satisfied for all $\rho \in \mathcal{F}_{\mathcal{P}}$, thus it is a convex problem, Rough & Shamma (2000); Wu (2001). In practice, this problem is set up by gridding the parameter space and solving the set of LMIs that hold on the subset of $\mathcal{F}_{\mathcal{P}}$, see Packard & Balas (1997). The LPV control is constructed by the Parameter Dependent Lyapunov Functions (PDLF) in which the conservatism of the control design is reduced.
To specify the scheduled performance weights for the LPV design the scheduling variables are defined through a lookup-table given on a suitable grid. The grid is determined by v, μ, R and ρ_f as follows: $v = [20, \ldots, 120]$kph, $\mu = [0.1, \ldots, 1.1]$ and $R = [0, R_a, R_b, 1]$. The scheduling parameter ρ_f, which is the fault information provided by the FDI filter, can be taken from interval $\rho_f = [0, 1]$. The zero value of ρ_f corresponds to the non-faulty operation and the value 1 to the full hydraulic actuator failure. The gridding reflects the qualitative changes of the performance weights, i.e. the scheduling variables. The robust stability and performance are guaranteed by the LPV design process, see Packard & Balas (1997); Wu (1995).

4. Design of tracking control

The starting point for the tracking control design of the active suspension actuator is a quarter car LPV model of the suspension system augmented with a nonlinear actuator dynamics. The actuator is a nonlinear switched system (bimodal system) where the switch is triggered by the sign of the damper velocity.
In Figure 4 a two-degree-of-freedom quarter-car model is shown. The body mass m_s represents the sprung mass, which corresponds to one of the corners of the vehicle, and the unsprung mass m_u represents the wheel at one corner. The parameters k_t, k_s, b_s are the tyre stiffness, the suspension stiffness, and the damping rate of the suspension, respectively. The control signal F is generated by the actuator. The disturbance d is caused by road irregularities.

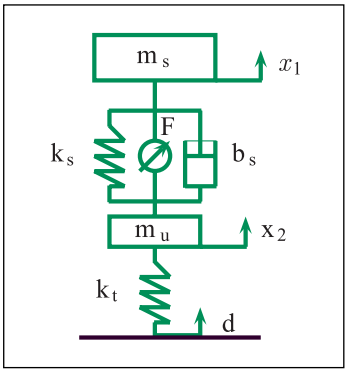


Fig. 4. Quarter-car model

The force equations of the quarter-car model are:

$$F_{m_s} = F_k + F_b - F, \tag{46}$$

$$F_{m_u} = -F_k - F_b - F_t + F. \tag{47}$$

The suspension damping force F_b is given by (9), the suspension spring force F_k is defined by (10) while the tire force F_t is given by (11) and F is the force of the actuator, respectively.
The state space representation of the quarter-car model can be formalized with the state vector $x = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix}^T$, where x_1 and x_2 denote the vertical displacement of the sprung mass and the unsprung mass, respectively, and x_3, x_4 denote their derivatives.

$$\dot{x}_3 = \frac{1}{m_s}(r_k(x_2 - x_1) + r_b(x_4 - x_3) + b_s^{nl}\rho_b\sqrt{\rho_b(x_4 - x_3)} - F), \tag{48}$$

$$\dot{x}_4 = \frac{1}{m_u}(-r_k(x_2 - x_1) - r_b(x_4 - x_3) - k_t(x_2 - d) - b_s^{nl}\rho_b\sqrt{\rho_b(x_4 - x_3)} + F), \tag{49}$$

where $r_b = b_s^l - b_s^{sym}\rho_b$ and $r_k = k_s^l + k_s^{nl}\rho_k$. Here $\rho_b = sgn(x_4 - x_3)$ and $\rho_k = (x_2 - x_1)^2$ are selected as scheduling variables.
An active actuator which generates the necessary force for the suspension system is a four-way valve-piston system, in general. Denoting by z the relative velocity one has $F = A_P P_L$, where A_P is the area of the piston and P_L is the pressure drop across the piston with respect to the front and rear suspensions. The derivative of P_L is given by

$$\dot{P}_L = -\beta P_L + \alpha A_P z + \gamma Q, \tag{50}$$

in which $Q = Q_0 x_v$ is the hydraulic load flow (with the notation $Q_0 = sgn(r)\sqrt{|r|}$ and $r = P_S - sgn(x_v)P_L$, moreover, α, β, γ are constants, P_S is the supply pressure and x_v is the displacement of the spool valve. The cylinder velocity acts as a coupling from the position output of the cylinder to the pressure differential across the piston. It is considered a feedback term, which has been analyzed by Alleyne & Liu (2000).
The displacement of the spool valve is controlled by the input to the servo-valve u:

$$\dot{x}_v = \frac{1}{\tau}(-x_v + u),$$

where τ is a time constant. Let x_5 and x_6 denote P_L and x_v, respectively. Then, the actuator model can be written separately as

$$\dot{x}_5 = -\beta x_5 + \alpha A_P z + \gamma Q_0(x_5, x_6)x_6, \tag{51}$$

$$\dot{x}_6 = -\frac{1}{\tau}x_6 + \frac{1}{\tau}u_a. \tag{52}$$

For both actuators it is hard to provide directly the command signals due to the high nonlinearities of these subsystems. Usually the controllers provide a force demand and in a second, postprocessing step the actual actuator commands are derived, see Alleyne & Hedrick (1995). The tracking control algorithm is derived by using a backstepping method applied for each of the modes.
In order to show the principle of the backstepping method the notations of van der Schaft (2000) are used. The model of the whole suspension and actuator system with zero disturbance are written in the following form

$$\dot{\zeta} = A\zeta + B\xi_1, \tag{53}$$

$$\dot{\xi}_1 = a_1(\zeta, \xi_1) + b_1(\xi_1)\xi_2, \tag{54}$$

$$\dot{\xi}_2 = a_2(\xi_2) + b_2 u, \tag{55}$$

where ζ is the state vector of the quarter-car suspension model, ξ_1 and ξ_2 are the state variables of the actuator dynamics and furthermore

$$a_1(\zeta,\xi_1) = -\beta\xi_1 + \alpha A_P z,$$
$$b_1(\xi_1) = \begin{cases} \gamma\sqrt{P_S - \xi_1}, & \xi_2 \geq 0 \\ \gamma\sqrt{P_S + \xi_1}, & \xi_2 < 0 \end{cases},$$
$$a_2(\xi_2) = -\frac{1}{\tau}\xi_2, \qquad b_2 = \frac{1}{\tau}.$$

Let us assume that there exists a smooth feedback function $K(\zeta)$ (possibly in LPV form) such that the closed loop system

$$\dot{\zeta} = A\zeta + BK(\zeta) \tag{56}$$

is asymptotically stable with control Lyapunov function $V(\zeta)$. The dynamics of (53) corresponds to the high-level suspension system (13). Therefore, the feedback function $K(\zeta)$ corresponds to the pressure demand required by the high-level control, i.e. $K(\zeta) = P_{L,dem} = \xi_{1,dem}$. In what follows $\xi_{2,dem} = \xi_{v,dem}$ denotes the demand of the spool valve displacement.
The backstepping design for the actuator subsystem can be performed in two steps. In the first step, let us consider $\xi_{2,dem}$ as a virtual input and $y_1 = \xi_1 - K(\zeta)$ as a virtual output. Since ξ_1 is not a manipulable input, we would like to construct a feedback that guarantees the tracking of $K(\zeta)$ with ξ_1. It is reasonable therefore to define the tracking error to be linear and stable, i.e., $\dot{y}_1 = -k_1 y_1$, $k_1 > 0$. Using (53)–(54) the desired time-function for $\xi_{2,dem}$ can be computed as a nonlinear feedback of the form

$$\xi_{2,dem} = \frac{1}{b_1(\xi_1)}[-a_1(\zeta,\xi_1) + \dot{K}(\zeta) - k_1(\xi_1 - K(\zeta))]. \tag{57}$$

In the second step, the desired input is u while the (virtual) output is defined as $y_2 = \xi_2 - \xi_{2,dem}$. For the tracking error, a stable linear dynamics is prescribed: $\dot{y}_2 = -k_2 y_2$, $k_2 > 0$. Using (53)–(55), we can now express the physically manipulable actuator input u as a function of ζ, ξ_1 and ξ_2 in the following form

$$u = \frac{1}{b_2}[-a_2(\xi_2) + \dot{\xi}_{2,dem}]. \tag{58}$$

By applying the above design, the closed loop system will be asymptotically stable with control Lyapunov function $S(\zeta) = V(\zeta) + \frac{1}{2}y_1^2 + \frac{1}{2}y_2^2$ Sepulchre et al. (1997). It is important to note that the obtained feedback law (58) is a state-dependent switching function because of the switching term $b_1(\xi_1)$ (see (56)) and it will be shown later that $\xi_{2,dem}$ can be approximated by a smooth function without affecting the validity of the method.
Since the actual feedback law generated by the LPV controller is a rather complicated function of the state variables, and we do not know the road excitation disturbances in advance, the above controller design procedure cannot be implemented in its original theoretical form. Therefore in what follows we will consider a more realistic assumption, when the reference for ξ_1 is computed by the high-level LPV controller, and for the trajectory tracking the time derivatives of the reference signals are computed numerically.
The reference for the pressure x_5, which is is denoted by $x_{5,dem}$, is computed by the high level LPV controller and for the trajectory tracking the time derivatives of the reference signals are computed numerically. The required tracking error dynamics is defined as

$$\dot{x}_5 - \dot{x}_{5,dem} = -k_1(x_5 - x_{5,dem}) \tag{59}$$

with a chosen positive constant parameter k_1.
Then the reference $x_{6,dem}$ is given by

$$x_{6,dem}\gamma\sqrt{P_S - sgn(x_{6,dem})x_5} = \Psi(x_5, x_{5,dem}, z), \tag{60}$$

i.e.

$$x_{6,dem} = \begin{cases} \frac{\Psi(x_5,x_{5,dem},z)}{\gamma\sqrt{P_S-x_5}}, & x_{6,dem} \geq 0 \\ \frac{\Psi(x_5,x_{5,dem},z)}{\gamma\sqrt{P_S+x_5}} & x_{6,dem} < 0 \end{cases}$$

with the notation $\Psi(x_{5,dem}, z) = \beta x_5 - \alpha A_p z + \dot{x}_{5,dem} - k_1(x_5 - x_{5,dem})$. Note that while Ψ is a smooth function, by taking the time derivative of both sides of (60) one has

$$\dot{x}_{6,dem} = \begin{cases} \frac{\dot{\Psi}(x_{5,dem},z)}{\gamma\sqrt{P_S-x_5}} + \frac{\dot{x}_{5,dem}x_{6,dem}}{2\gamma|P_S-x_5|}, & x_{6,dem} > 0 \\ \frac{\dot{\Psi}(x_{5,dem},z)}{\gamma\sqrt{P_S+x_5}} - \frac{\dot{x}_{5,dem}x_{6,dem}}{2\gamma|P_S+x_5|} & x_{6,dem} < 0 \end{cases}$$

i.e., $x_{6,dem}$ in general is not smooth – it is continuous but not differentiable. However it can be arbitrary approximated by a smooth function $\bar{x}_{6,dem}$ for which tacking the tracking error dynamics

$$\dot{x}_6 - \dot{\bar{x}}_{6,dem} = -k_2(x_6 - \bar{x}_{6,dem}). \tag{61}$$

with a chosen positive constant parameter k_2 is meaningful. Moreover this $\bar{x}_{6,dem}$ can be chosen such that $\bar{x}_{6,dem} \neq x_{6,dem}$ only on a small neighborhood of the origin (the discontinuity point of $\dot{x}_{6,dem}$), e.g. by taking an expression on $|\Psi| \leq \epsilon$ linear in Ψ, i.e. $\bar{x}_{6,dem} = \beta\Psi$ with a suitable β. Since $x_{6,dem}$ should satisfy (60) this choice does not affect the values of the desired x_5 considerably.
Finally, the following expression for the physical input is deduced:

$$u_a = x_6 + \tau\dot{\bar{x}}_{6,dem} - \tau k_2(x_6 - x_{6,dem}). \tag{62}$$

In order to practically implement the control law, we need to compute the time derivatives of $x_{5,dem}$ and $x_{6,dem}$, which can be done in a number of ways depending on the measurement noise conditions and the required precision, for details see Gáspár et al. (2008).

5. Design of the FDI filter

Significant research results have been published for the general FDI problem and several methods have been proposed, e.g. the parity space approach, the multiple model method, detection filter design using a geometric approach, or the dynamic inversion based detection, see Massoumnia (1986); Gertler (1997); Szigeti et al. (2001). Most of the design approaches refer to linear, time-invariant (LTI) systems, but references to some nonlinear cases are also found in the literature, see Stoustrup & Niemann (1998); Chen & Patton (1999). An $\mathcal{H}_\infty$ approach to design a fault detection and isolation gain-scheduled filter for *LPV* systems was presented by Abdalla et al. (2001); Bokor & Balas (2004). There are also numerous papers dealing with the design of reconfigurable controls, which include the design of FDI filters, the design of reconfigurable controllers and the design of reconfiguration mechanisms. Applications of reconfigurable control systems are found in different fields, see e.g. Fischer & Isermann (2004); Kanev & Verhaegen (2000).

Possible faults of the actuators (loss of effectiveness) can be detected by reconstructing the actual suspension forces. Having measured the signals $y_1 = \dot{x}_3$,$y_2 = \dot{x}_4$ and $y_3 = x_2 - x_1$ an inversion based detection filter is proposed, Balas et al. (2004); Szabó et al. (2003). In the construction of the filter the first step is to express F from (48) and in these expression we plug in the known values y_i:

$$F = |z| + b_s^{nl}\rho_b\sqrt{|z|} + r_k y_3 - m_s y_1. \tag{63}$$

In this expression the value of the relative velocity z is not measured. The road disturbance is an unknown input signal but from the equations (48), (49) one has

$$m_s\dot{x}_3 + m_u\dot{x}_4 = -k_t(x_2 - d). \tag{64}$$

By plugging back the obtained expressions in the original equations one has the system $\dot{x}_3$ $= \frac{r_k}{m_s}(x_2 - x_1)$ - $\frac{r_k}{m_s}y_3 + y_1$ and $\dot{x}_4$ =- $\frac{r_k}{m_u}(x_2 - x_1)$ + $\frac{r_k}{m_u}y_3 + y_2$, where the relative velocity is not measured. The resulting LPV system

$$\dot{z} = -r_k m_e z + r_k m_e y_3 + y_2 - y_1, \tag{65}$$

with $m_e = \frac{m_u + m_s}{m_u m_s}$ will be observable.
For active actuators, since the real actuators might present a saturation effect, in addition to compare the reconstructed forces with the force demands provided by the robust LPV controllers it is necessary to check, if the actual forces are lower then those corresponding to the saturation level of the actuators.
To obtain the final fault detection filter equations (51) and (52) are used as:

$$\dot{\tilde{x}}_5 = -\beta\tilde{x}_5 + \alpha A_P\hat{z} + \gamma Q_{0,nom}(\hat{F})\tilde{x}_6, \tag{66}$$

$$\dot{\tilde{x}}_6 = -\frac{1}{\tau_{nom}}\tilde{x}_6 + \frac{1}{\tau_{nom}}u_a, \tag{67}$$

where $\hat{z}$ and $\hat{F}$ are the estimated damper velocity and damper force values, respectively. A possible actuator fault affects the terms Q_0 through a modified value of P_s and the time constant τ, respectively. The nominal values of these parameters (i.e. for the fault free case) are denoted by the subscript *nom*.
For the fault free case one should have $e_5 = x_5 - \tilde{x}_5 \approx 0$ and $e_6 = x_6 - \tilde{x}_6 \approx 0$, respectively. Since the initial conditions are not known, an observer need to be constructed for (65) and (66), (67) respectively, to test these conditions. For a Leuenberger–type observer a design method for bimodal systems was reported in Juloski et al. (2007). For this case, however, the nonsmooth term Q_0 – which in turn makes the system to be bimodal – is considered as scheduling variable. Hence a more conventional LPV observer can be constructed.
For a LPV system that depends affinely on the scheduling variables an LPV observer gain can be designed using LMI techniques: let us recall that an LPV system is said to be quadratically stable if there exist a matrix $P = P^T > 0$ such that $A(\rho)^T P + PA(\rho) < 0$ for all the parameters ρ. A necessary and sufficient condition for a system to be quadratically stable is that this condition holds for all the corner points of the parameter space, i.e., one can obtain a finite system of linear matrix inequalities (*LMI*s) that have to be fulfilled for $A(\rho)$ with a suitable positive definite matrix P, see Gahinet (1996).
In order to obtain a quadratically stable observer the *LMI*

$$A_o^T(\rho)P + PA_o(\rho) < 0 \tag{68}$$

must hold for suitable $K(\rho)$ and $P = P^T > 0$, with $A_o = A + KC$. By introducing the auxiliary variable $L(\rho) = PK(\rho)$, one has to solve the following set of *LMIs* on the corner points of the parameter space:

$$A(\rho)^T P + PA(\rho) - C^T L(\rho)^T - L(\rho)C < 0. \tag{69}$$

By solving these LMIs a suitable observer gain is obtained:

$$K(\rho) = P^{-1} L(\rho). \tag{70}$$

If $e = ||e_5||_2 + ||e_6||_2$ is greater than a given threshold, then a fault must be present in the system and a fault signal is emitted to the higher level controller, used in the controller reconfiguration process. The threshold level influences the fault-detection delay, i.e. high threshold level corresponds to increased delay. However, due to disturbances, sensor noises and the modeling uncertainties this level cannot be arbitrarily small and it is determined using engineering knowledge.

6. Simulation example

In this section the operation of the integrated control is presented and analyzed through simulation examples.

In the first example the operation of the two-level controller is demonstrated. The controller, which combines a high-level LPV controller and a low-level nonlinear controller, is built in Matlab/Simulink software environment.

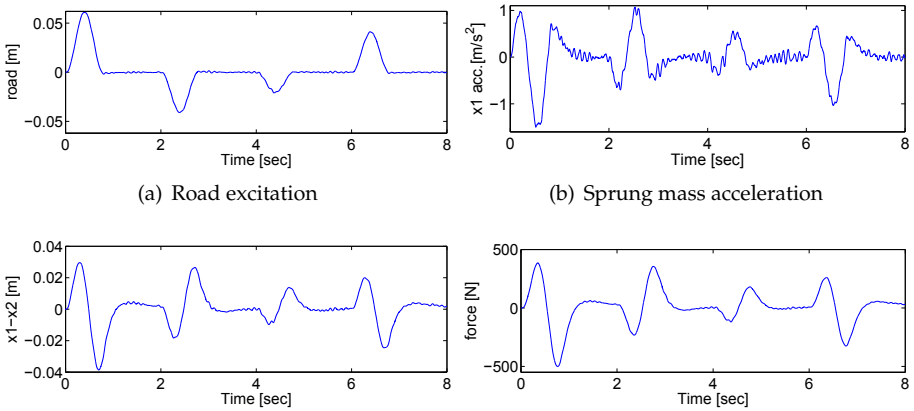
In the simulation example an upper-level controller is designed based on the *LPV* method, which generates a required control force. The controlled systems are tested on a bad-quality road, on which bumps of four different heights disturb the motion of the vehicle: the height of the bumps are 6 *cm*, 4 *cm*, 2 *cm* and 4 *cm*, respectively. Between the bumps there are velocity-dependent stochastic road excitations. The time responses of the road excitation, the heave acceleration, the relative displacement and the control force in the front and left-hand-side are illustrated in Figure 5. The bumps with extremely large amplitude cause large acceleration of the sprung mass and large relative displacement between the two masses.

Thanks to the controller the effects of the road disturbances on the performances are acceptable since the values of the performance signals tend to zero in a short time period. The suspension problem is solved by the force defined by the controller in the upper-level.

Then the low-level controller is applied in order to track the designed force. The operation of the force-tracking controller based on the backstepping method is illustrated in Figure 6. In the control design the parameters are selected as $k_1 = 20$ and $k_2 = 20$. In the simulation example it is assumed that the sampling time of the measured signals is selected $T_s = 0.01$ *sec*, which corresponds to practice.

The illustrated signals are the pressure drop across the piston, the displacement of the spool valve, the control input, the achieved force and the RMS of the force error. The achieved force generated by the actuator tends to the required force. The RMS of the force error, see Figure 7, shows that the generated force approximates the required force with high precision.

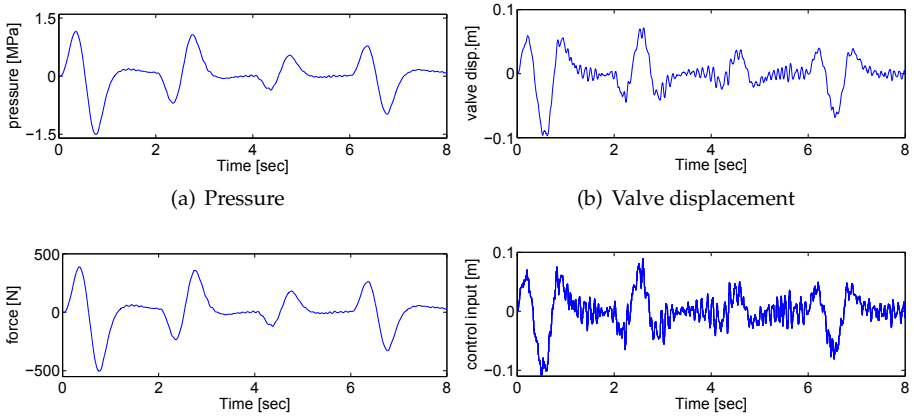
The second example illustrates the operation of the FDI filter applied to an active suspension system. The dashed red line presents the required force designed by the control system. The current force must be calculated by using the measured signals. A filter is used to calculate the current force by using an inversion method and the measured signals, i.e. the accelerations



(a) Road excitation (b) Sprung mass acceleration

(c) Relative displacement (d) Force required by upper level controller

Fig. 5. Control input required by the upper-level controller



(a) Pressure (b) Valve displacement

(c) Force generated by low level controller (d) Control input

Fig. 6. Analysis of the tracking properties using the backstepping method

of the sprung mass and the unsprung mass, and the relative displacement between the two masses.

The reconstructed force is illustrated by the solid blue line in the upper plot of Figure 8. The force is compared with the force produced by a fault free suspension system (dashed line). The FDI filter also gives the signals depicted in blue in the lower plot of Figure 8, while the red signal is the chosen threshold level expressed in a given percent of the desired force. Since the obtained error level will be greater than this threshold, a fault signal is emitted indicating a faulty actuator.

In the third example the operation of the fault-tolerant integrated control that uses the designed FDI filter is illustrated. The vehicle performs a cornering maneuver with 70 km/h.

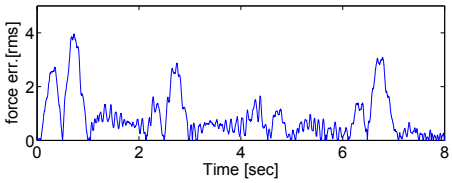


Fig. 7. Force error (RMS)

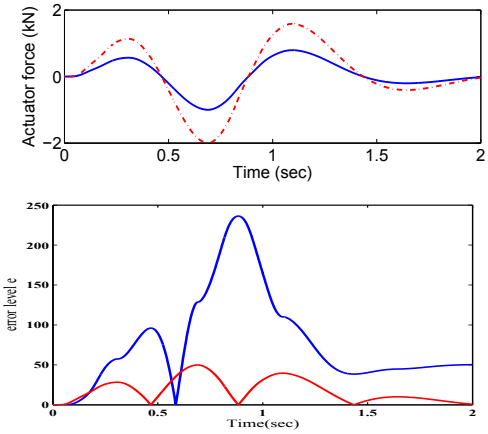


Fig. 8. The result of the FDI procedure

velocity. During the cornering maneuver the lateral acceleration increases and thus the roll angle of unsprung masses also increases. The time response of the steering angle, the lateral acceleration, the forward velocity and the normalized lateral load transfer at the rear are depicted in Figure 9.

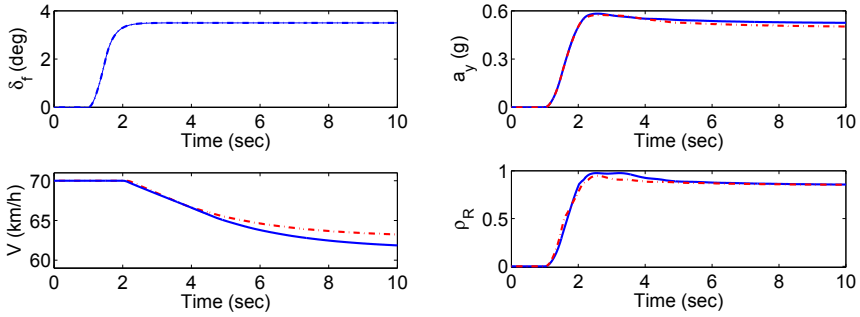


Fig. 9. Time responses of the control system

Since the monitoring scheduling variable, i.e., the normalized lateral load transfer ρ_R increases the suspension system generate stabilizing roll moment to balance the overturning moment. However, the normalized lateral load transfer also exceeds the predefined critical value R_a

and the brake generated a force with which the direction of the vehicle slightly modified and consequently the effect of the lateral force reduces. Figure 10 shows the control signals, i.e. the braking force at the rear and all the suspension forces.
Then it is assumed that an actuator failure in the suspension system has already been detected at the front and right. The time response of the control signals are also depicted in Figure 10. The solid blue line illustrates the fault operation and the dashed red line illustrates the fault-free case.

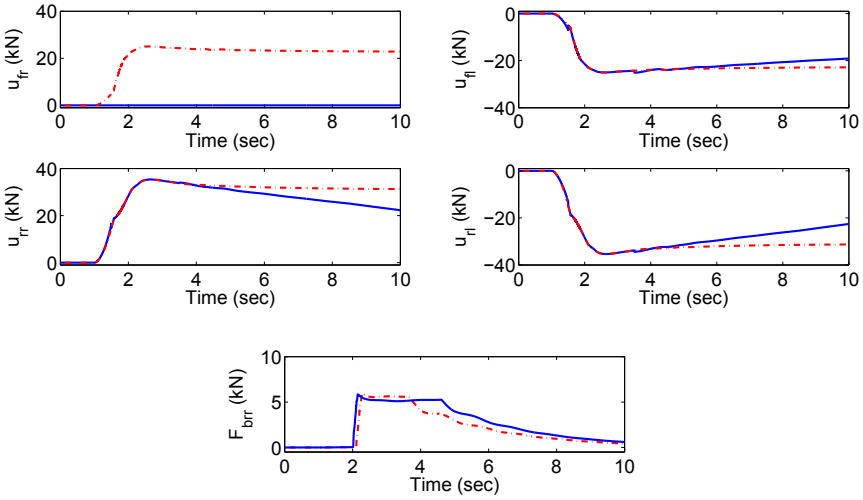


Fig. 10. Time responses of the control signals

It is observed that the normalized load transfer increases due to the reduced power of the actuators. According to the detected actuator fault the brake is activated at a smaller value of the critical normalized load transfer. Moreover, the duration of the required brake force is longer in the case of a suspension fault. Because of the braking action the suspension system generates the same forces (except in the fault component) as they are in the fault-free case.
In the fourth example the selection of R_a and R_b regarding the activation of the brake is critical. If the brake is activated at a large R_a the probability of rollover increases. If the value R_a was small, the brake would be activated very frequently. In case of a fault the selection of ρ_D also has an important role. Finally, we shall examine the effects of varying the design parameter R_a on the controlled system. In Figure 11 the peak lateral acceleration against forward velocity is plotted during a vehicle maneuver. R_b is fixed at 0.95 and R_a varies. The dash-dot, dashed and solid lines correspond to $R_a = 0.7$, $R_a = 0.8$ and $R_a = 0.9$ respectively. With $R_a = 0.7$ in the controlled system there is a gradual brake control, whereas when $R_a = 0.9$, the brake system is not used until the normalized load transfer ρ_R equals 0.9, and the response of the yaw-roll model is the same as when only active suspensions are used. Thus the design parameters R_a and R_b can be used to shape the nonlinear response characteristics of the controlled system.
In a non-faulty case, which means that suspension system is working well, it would be preferable to choose R_a large. This corresponds to an active brake system that is not used for most of the time and activated very rapidly when the normalized load transfer exceeds the critical

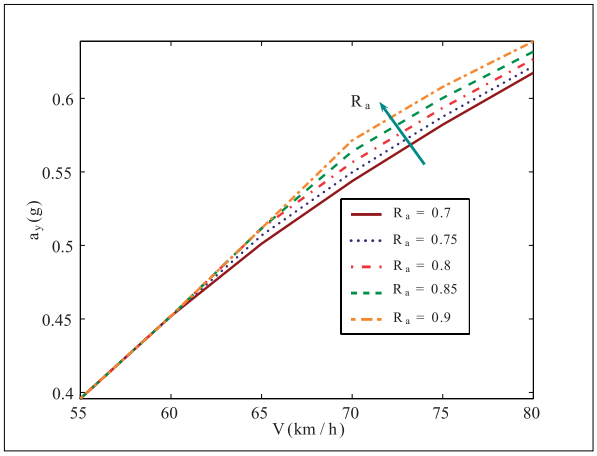


Fig. 11. Effect of parameter R_a on lateral acceleration

value determined by R_a. However, this would result in a large lateral acceleration until the critical R_a is reached. This would be a small price for the stability of roll motion. Because until the critical R_a has been reached only the active suspensions, which do not affect directly the roll dynamics of the vehicle, are used.
On the other hand, if a hydraulic actuator fault occurs in the system it would be preferable to choose R_a small. This corresponds to a combined control where the range of operation of the brake system is extended and the wheels are decelerated gradually rather than rapidly if the normalized load transfer has reached R_a. It is assumed that the actuator fault can occur as a loss of effectiveness, i.e. its power is reduced by some percent. It means that both control inputs are able to work simultaneously but the hydraulic actuator does not have maximum performance. It is a reasonable assumption in many cases because the occurrence of the failure indicates an effectiveness failure at an early stage. As a consequence, the design parameter R_a can be chosen as a scheduling parameter based on the fault information.

7. Conclusions

In this paper an application of the Linear Parameter Varying method for the design of integrated vehicle control systems has been presented, in which several active components has been used in co-operation. In the control design besides performance specifications and uncertainties, the fault information has been taken into consideration. By monitoring suitable scheduling parameters in the LPV control, the reconfiguration of the control systems can be achieved, conflict between performance demands can be avoided and faults (loss in effectiveness) can be handled.
In the proposed scheme if a fault occurs in the active suspension system and it is detected by the FDI filter, the active brake assumes the role of the active suspension to enhance rollover prevention. A weighting strategy is applied in the closed-loop interconnection structure, in which the normalized lateral load transfer and the residual output of the FDI filter play an important role. A tracking controller and an FDI filter has been designed that provides the reference signal for the low-level actuator design and it also constitutes the supervisor con-

troller for the reconfiguration. By using the LPV method the designed controller guarantees the desired stability and performance demands of the closed–loop system.

8. Acknowledgements

This work is supported by the Hungarian National Office for Research and Technology through grants TECH_08_2/2-2008-0088 is gratefully acknowledged.
The effort was sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under the grant number FA8655-08-1-3016. The U.S Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon.

9. References

Abdalla, M., Nobrega, E. & Grigoriadis, K. (2001). Fault detection and isolation filter design for linear parameter varying systems, *Proceedings of the American Control Conference 2001*, Vol. 5, IEEE, Arlington, VA, USA, pp. 3890–3895.

Alleyne, A. & Hedrick, J. (1995). Nonlinear adaptive control of active suspensions, *IEEE Transactions on Control Systems Technology* **3**(1): 94–101.

Alleyne, A. & Liu, R. (2000). A simplified approach to force control for electro-hydraulic systems, *Control Engineering Practice* **8**(12): 1347–1356.

Balas, G., Bokor, J. & Szabó, Z. (2004). Tracking of continuous LPV systems using dynamic inversion, *Proceedings of the 43rd IEEE Conference on Decision and Control 2004*, IEEE, San Diego, CA, USA, pp. 2929–2933.

Becker, G. & Packard, A. (1994). Robust performance of linear parametrically varying systems using parametrically-dependent linear feedback, *Systems & Control Letters* **23**(3): 205–215.

Bokor, J. & Balas, G. (2004). Detection filter design for LPV systems - a geometric approach, *Automatica* **40**(3): 511–518.

Chen, B. & Peng, H. (2001). Differential-braking-based rollover prevention for sport utility vehicles with human-in-the-loop evaluations, *Vehicle System Dynamics* **36**(4–5): 359–389.

Chen, J. & Patton, R. (1999). *Robust Model based Fault Diagnosis for Dynamic Systems*, Kluwer, Boston/Dordrecht/London.

Fischer, D. & Isermann, R. (2004). Mechatronic semi-active and active vehicle suspensions, *Control Engineering Practice* **12**(11): 1353–1367.

Gahinet, P. (1996). Explicit controller formulas for LMI-based $\mathcal{H}_\infty$ synthesis, *Automatica* **32**(7): 1007–1014.

Gertler, J. (1997). Fault detection and isolation using parity relations, *Control Engineering Practice* **5**(5): 653–661.

Gillespie, T. (1992). *Fundamentals of vehicle dynamics*, Society of Automotive Engineers Inc.

Gáspár, P., Szabó, Z. & Bokor, J. (2006). Side force coefficient estimation for the design of active brake control, *Proceedings of the American Control Conference 2006*, IEEE, Minneapolis, MN, USA, pp. 2927–2932.

Gáspár, P., Szabó, Z., Szederkényi, G. & Bokor, J. (2008). Two-level controller design for an active suspension system, *Proceedings of the 16th Mediterranean Conference on Control and Automation 2008*, IEEE, Ajaccio-Corsica, France, pp. 232 – 237.

Juloski, A. L., Heemels, W. P. M. H. & Weiland, S. (2007). Observer design for a class of piecewise linear systems, *International Journal of Robust and Nonlinear Control Engineering Practice* **17**(15): 1387–1404.

Kanev, S. & Verhaegen, M. (2000). Controller reconfiguration for non-linear systems, *Control Engineering Practice* **8**(11): 1223–1235.

Leith, D. & Leithead, W. (2000). Survey of gain-scheduling analysis and design, *International Journal of Control* **73**(11): 1001–1025.

Massoumnia, M. A. (1986). A geometric approach to the synthesis of failure detection filters, *IEEE Transactions on Automatic Control* **31**(9): 839–846.

Packard, A. & Balas, G. (1997). Theory and application of linear parameter varying control techniques, *Proceedings of the American Control Conference 1997, Workshop I*, Albuquerque, New Mexico.

Palkovics, L., Semsey, A. & Gerum, E. (1999). Roll-over prevention system for commercial vehicles, *Vehicle System Dynamics* **32**(4–5): 285–297.

Rough, W. & Shamma, J. (2000). Research on gain scheduling, *Automatica* **36**(11): 1401–1425.

Sampson, D. & Cebon, D. (2003). Active roll control of single unit heavy road vehicles, *Vehicle System Dynamics* **40**(4): 229–270.

Sepulchre, R., Jankovic, M. & Kokotovic, P. (1997). *Constructive Nonlinear Control*, Springer-Verlag.

Sharp, R. & Crolla, D. (1987). Road vehicle suspension system design: A review, *Vehicle System Dynamics* **16**(3): 167–192.

Stoustrup, J. & Niemann, H. (1998). Fault detection for nonlinear systems - a standard problem approach, *Proceedings of the 37th IEEE Conference on Decision and Control 1998*, Vol. 1, IEEE, Tampa, FL, USA, pp. 96–101.

Szabó, Z., Bokor, J. & Balas, G. (2003). Inversion of LPV systems and its application to fault detection, *Proceedings of the 5th IFAC Symposium on Fault Detection Supervision and Safety for Technical Processes (SAFEPROCESS) 2003*, Washington, DC, USA, pp. 235–240.

Szigeti, F., Vera, C., Bokor, J. & Edelmayer, A. (2001). Inversion based fault detection and isolation, *Proceedings of the 40th IEEE Conference on Decision and Control 2001*, Vol. 2, IEEE, Orlando, FL, USA, pp. 1005–1010.

van der Schaft, A. J. (2000). *L2-Gain and Passivity Techniques in Nonlinear Control*, Springer-Verlag, Berlin.

Wu, F. (1995). Control of linear parameter varying systems, *PhD Thesis, Mechanical Engineering, University of California at Berkeley* .

Wu, F. (2001). A generalized LPV system analysis and control synthesis framework, *International Journal of Control* **74**(7): 745–759.

6

Effect of Switchover Time in Cyclically Switched Systems

Khurram Aziz
School of Electrical Engineering and Computer Science,
National University of Sciences and Technology
Pakistan

1. Introduction

Cyclic service queueing systems have a broad range of applications in communication systems. From legacy systems like the slotted ring networks and switching systems, to more recent ones like optical burst assembly, Ethernet over SDH/SONET mapping and traffic aggregation at the edge nodes, all may employ cyclic service as a means of providing fairness to incoming traffic. This would require the server to switch to the next traffic stream after serving one. This service can be exhaustive, in which all the packets in the queue are served before the server switches to the next queue, or non-exhaustive, in which the server serves just one packet (or in case of batch service, a group of packets) before switching to the next queue.
Most of the study on systems with cyclic service has been performed on queues of unlimited size. Real systems always have finite buffers. In order to analyze real systems, we need to model queues with finite capacity. The analysis of such systems is among the most complicated as it is very difficult to obtain closed-form solutions to systems with finite capacity.
An important parameter in cyclic service queueing systems with finite capacity is the switchover time, which is the time taken for the server to switch to a different queue after a service completion. This is especially true for non-exhaustive cyclic service systems, in which the server has to switch to the next queue after serving each packet. The switchover time is usually very small as compared to the service time, and is generally ignored during analysis. In such cases, the edge node can be modeled as a server, serving the various access nodes - that can be modeled as queues - in a cyclic manner. Hence, we assume that on finding an empty queue, the server will go to the next queue with a switchover rate of, say ε, but if the queue is not empty, we ignore the switchover time and assume that the server will switch to the next queue with rate μ after serving one packet in the queue.
While this generally led to quite accurate results in the past due to a large difference in ratios between the service and switchover times, this might not be the case today as optical communication systems are getting faster and faster. Thus the switchover time cannot always be safely ignored as smaller differences between switchover times and service times may introduce significant differences in the results. In order to analyze such systems, the switchover process can be modeled as another phase in the service process.
The focus of this chapter is on the analysis of non-exhaustive cyclic service systems with finite capacity using state space modeling technique. A brief summary on the work done to date, in cyclic service systems is presented in Section 2, while some applications of such systems

are discussed in Section 3. Analytical models of systems in which the switchover times can be ignored during service are presented in Section 4, in which we start from a simple two-queue system and generalize for an n-queue system. Analytical study of edge nodes that employ non-exhaustive cyclic service to serve various incoming streams as a two stage process (serving and switchover) in which switchover times are not ignored during service is presented in Section 5, followed by a detailed comparison of systems with and without switchover times in Section 6. Scenarios in which switchover cannot be ignored are also discussed in this section. Finally, the results are summarized in Section 7.

2. Related work

The study on cyclic service queueing systems is quite extensive. It would thus be helpful if these systems can be categorized. Several types of classifications have been presented in the literature, with the most recent being the survey by Vishnevskii and Semenova (Vishnevskii & Semenova, 2006). The classification presented here, however, is based on the most widely used parameters and related work in those categories is then presented.

2.1 Categorization of cyclic service queueing models

In a cyclic service queueing system, two or more queues are associated with the same server, which scans different queues in a round-robin manner and serves the queue if a packet is present. This service can be of three types – exhaustive, non-exhaustive and gated. In an exhaustive service model, the server switches over to the next queue only after completing service for all the customers in the queue. This also includes any new customers that may arrive during this time. On the other hand, in a gated system, service is provided to only those customers that were present in the queue when the server arrives to that queue. A limited, or non-exhaustive service is one in which a fixed number of customers – typically one – are serviced by the server during one visit. Usually, the exhaustive service is considered more efficient in terms of the waiting time of the customers than the gated service, which in turn is considered more efficient than the non-exhaustive service. Hence, in case of the exhaustive and gated service policies, queues with a large number of customers get more attention than those with a small number of customers, resulting in a less fair service as compared to the non-exhaustive service policy. So depending on the definition of "fairness", the non-exhaustive service policy is the fairest. This is especially true for communication systems, as different queues usually represent different traffic streams and it is undesirable to prefer one stream to another if their priorities are equal.

Another important consideration in such systems is the switchover time which is the time taken by the server to move to the next queue, after finishing service in the current queue. The switchover time is usually quite small as compared to the service time and is ignored in most studies. However, this can cause large differences in results especially if the switchover rate is not large as compared to the service rate.

In addition to the finite switchover rate, another issue is the size of the queues. Most of the studies on systems with cyclic service have been performed on queues of unlimited size. Real systems always have finite buffers. In order to analyze real systems, queues with finite capacity need to be modelled. An important feature of such systems is blocking, which happens when the queue becomes full and any subsequent arrivals are lost.

The cyclic service queueing models can thus be mainly categorized in the following different ways:

- Service discipline – exhaustive, gated and non-exhaustive.

- Switchover times – zero and non-zero.
- Buffer capacity – infinite, finite and single buffer.

2.2 Cyclic service systems with infinite buffers

Cyclic service queueing systems have been extensively studied in the literature. The first study on the cyclic polling systems available is the patrolling machine repairman model (Mack et al., 1957) where a single repairman visits a sequence of machines in cyclic order, inspecting them and repairing them when failure has occurred. The first study on cyclic polling models relating to communication networks was in the early 1970s to model the time-sharing computer systems. Since then, there has been an extensive research in this area, especially since the range of applications in which cyclic polling models can be used is very broad.

Leibowitz (Leibowitz, 1961) was among the first to study an approximate solution for symmetrically loaded cyclic polling system with gated service and constant switchover time. Cooper and Murray (Coooper & Murray, 1969; Cooper, 1970) analyzed exhaustive and gated service systems using an imbedded Markov chain technique for zero switchover time. Eisenberg (Eisenberg, 1971) studied a two-queue system with general switchover time, while Eisenberg (Eisenberg, 1972) and Hashida (Hashida, 1972) generalized the results of Cooper and Murray for non-zero switchover times. Bux and Truong (Bux & Truong, 1983) provided a simple approximation analysis for an arbitrary number of queues, constant switchover time and exhaustive service discipline. Lee (Lee, 1996) studied a two-queue model where the server serves customers in one queue according to an exhaustive discipline and the other queue according to a limited discipline, while Boxma (Boxma, 2002) studied a combination of exhaustive and limited disciplines in the two queues along with a patient server, which waits for a certain time in case there are no customers present in one of the queues.

For non-exhaustive cyclic service and general switchover times, Kuehn (Kuehn, 1979) developed an approximation technique based on the concept of conditional cycle times and derived a stability criteria for the general case of $GI/G/1$ systems with a cyclic priority service. Boxma (Boxma, 1989) related the amount of work in a polling system with switchover times to the amount of work in the same polling system without switchover times, leading to several studies on this relationship, notably by Cooper et al. (Cooper et al., 1996), Fuhrmann (Fuhrmann, 1992), Srinivasan et al. (Srinivasan et al., 1995), and Borst and Boxma (Borst & Boxma, 1997). An important question is that how large should the switchover rate be as compared to the service rate, so that it can be safely ignored. The answer is not simple and this study will attempt to answer this question in relation to the cyclic service queueing models with finite buffers and non-exhaustive service in later sections.

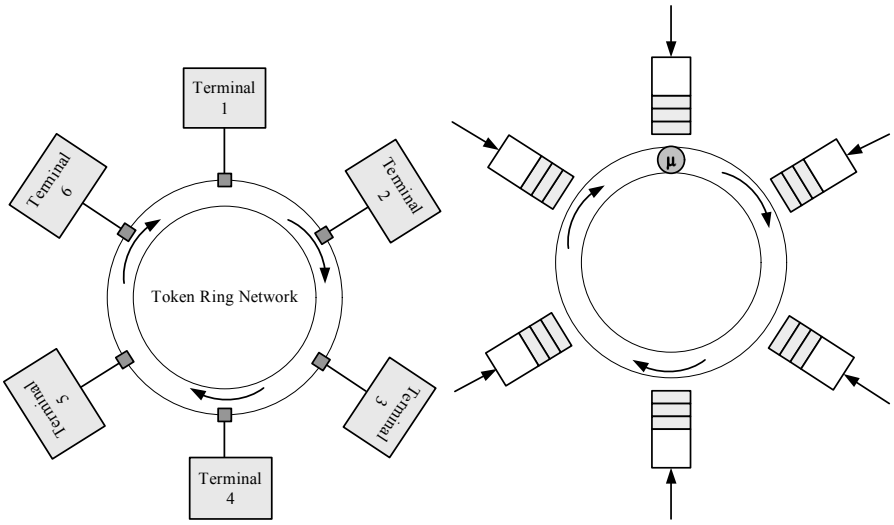
2.3 Cyclic service systems with finite buffers

While the study of cyclic service systems with infinite buffer capacity has been very extensive and closed form solutions for several such systems with exhaustive service discipline exist, the study of cyclic service systems with finite capacity queues and non-exhaustive service discipline is rather limited in the literature. Single buffer systems have been studied by Chung and Jung (Chung & Jung, 1994), and Takine et al., (Takine et al., 1986; 1987; 1990). Magalhaes et al., (Magalhaes et al., 1998) present a distribution function for the interval between the instant when the customers leave each queue, in a two-queue $M/M/1/1$ polling system. Titenko (Titenko, 1984) established formulae for the calculation of the moments of any order of the waiting times for single-buffer queues. Takagi (Takagi, 1992) presented the Laplace-Stieltjes transform (LST) of the cycle time for an exhaustive service, $M/G/1/n$ polling

system. A virtual buffer scheme for customers entering the system when the queue is full is suggested by Jung (Jung & Un, 1994). Tran-Gia and Raith have several important studies in this area. In (Tran-Gia & Raith, 1985a;b), a non-exhaustive cyclic queueing systems with finite buffers is analyzed based on the imbedded Markov chain approach in conjunction with a two-moment approximation for the cycle time. In (Tran-Gia, 1992), the stationary probability distributions of the number of waiting customers at polling instants as well as arbitrary instants for a $GI/G/1/n$ polling system with a 1-limited service discipline is obtained using discrete time analysis. Onvural and Perros (Onvural & Perros, 1989) present an approximation method for obtaining the throughput of cyclic queueing networks with blocking as a function of the number of customers. A polling system with Munit capacity queues and one infinite capacity queue with exhaustive service is described in (Takine et al., 1990).
The work on polling systems has been well summarized by Takagi in various papers. In (Takagi, 1986), all the results available till 1986 were organized, while an up-to-date summary on polling systems was presented in (Takagi, 1988). This survey was updated twice in 1990 (Takagi, 1990) for all work until 1989 and 1997 (Takagi, 1997) for the advances made after his previous update between the years 1990 to 1994. A more recent survey by Vishnevskii and Semenova (Vishnevskii & Semenova, 2006) covers various polling models, including the cyclic service models with finite service in great detail. It is clear, however, that accurate and generalized results for cyclic service finite queueing models are still not available. Some authors have provided a few closed form solutions for some specific models, but most of the time, these are approximate solutions, mostly for single buffer systems.

3. Applications of cyclic service queueing systems

Polling models with cyclic service can be used in a wide range of applications, from computer communications to robotics, production, manufacturing, and transportation. In computer communications, the queueing model with cyclic service was first used in the analysis of time-sharing computer systems in the early 1970's. In the 1980's, the token passing systems such as the token ring and token bus, as well as other demand-based channel access schemes in local area networks, such as the one shown in Figure 1, were analysed using such queueing systems with cyclic service.
From legacy systems like the slotted ring networks and switching systems, to more recent systems like wireless networks, optical burst assembly, Ethernet over SDH/SONET mapping and traffic aggregation at the edge nodes, and all may employ cyclic service as a means of providing fairness to incoming traffic. Queueing systems with cyclic service are extensively used especially at the edge nodes to provide fairness to the different flows that arrive at the node. One such example is the fair queueing system proposed by Nagel (Nagle, 1987). Another example is the mapping of Ethernet over SDH/SONET, as shown in Figure 2. Cyclic service can also be employed by the burst assembler in an optical burst switching node as shown in Figure 3.
Ibe and Trivedi (Ibe & Trivedi, 1990) propose the use of stochastic Petri Net models for obtaining the performance measures of a finite buffer polling system using the exhaustive, gated and limited service disciplines. Choi (Choi, 2004) proposes a cyclic polling based algorithm for differentiated class of services in Ethernet passive optical networks. Takagi highlights three classical but instructive applications of polling models to the performance evaluation of communication networks in (Takagi, 2000). The three applications discussed are the half-duplex transmission for an inquiry system, the polling data link control and the token ring network. Bruneel and Kim (Bruneel & Kim, 1993), Grillo (Grillo, 1990), and Levy and Sidi (Levy & Sidi,



(a) Token ring network (b) Equivalent queueing model

Fig. 1. Token ring network and its equivalent queueing model

1990) analyze several applications examples of communication networks, including the ATM systems that employ the cyclic polling systems. In the late nineties, rapid development of the broadband wireless transmission networks prompted several studies of the polling system models in this area, especially by Ziouva and Antonakopoulos (Ziouva & Antonakopoulos, 2002/2007; 2003), and Vishnevskii (Vishnevsky et al., 1999; 2004). Miorandi et al., (Miorandi et al., 2004) performed an interesting study on the performance evaluation of the Bluetooth polling schemes.

The focus in this chapter is to study the basic polling models that employ non-exhaustive cyclic service and finite queues, independent of the communication system involved, and study the effect of switchover time on these systems.

4. Systems with zero switchover times

In this section, cyclic service queueing systems that ignore the switchover times during service are studied. Typically, a server spends some time serving a customer and then switches over to the next queue. The time taken for the server from the completion of service in one queue to the commencement of service in the next queue is known as the switchover time. This switchover takes a small amount of time as compared to the service time and is usually ignored. The assumption here is that the switchover times in such systems will be very small as compared to the service times and when ignored, they will not have a considerable effect on the overall system performance.

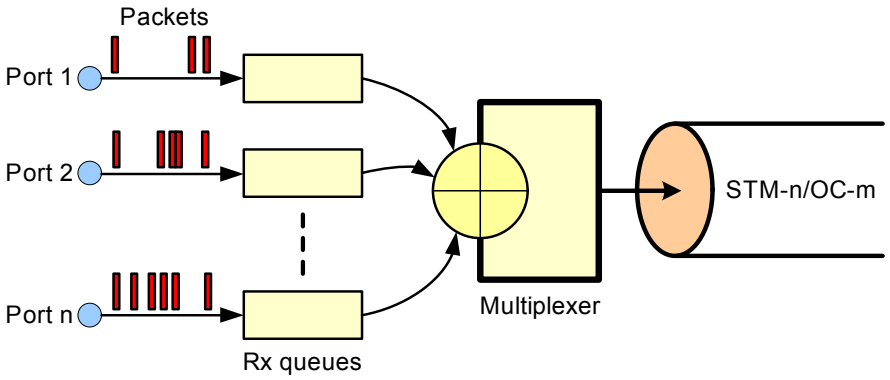


Fig. 2. Mapping Ethernet over SDH/SONET in an edge node

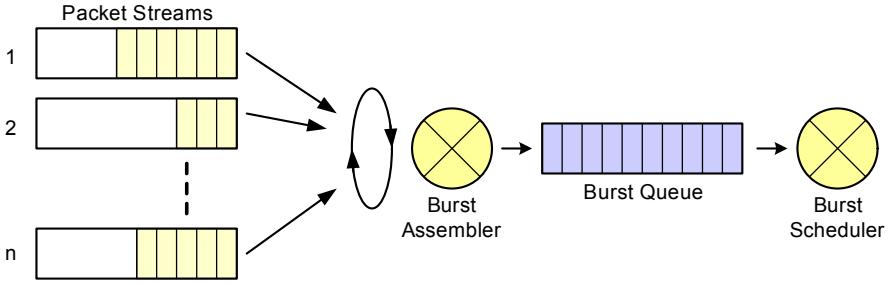


Fig. 3. Burst assembly in an OBS edge node using cyclic service

4.1 Model description

Cyclic service systems can be modelled as shown in Figure 4, which shows N queues, each of size s_i $(i = 1,\ldots,N)$, being served in a round-robin manner by a server with an exponentially distributed service rate of mean μ. The arrival rate to each queue is also exponentially distributed with mean λ_i $(i = 1,\ldots,N)$. The average time taken by the server to switch over from one queue to the next is given by $1/\varepsilon$ where ε is the mean switchover rate.

At each scanning epoch, the server processes one packet in the queue if there is at least one packet waiting. In case there is no waiting packet in the queue, the server switches over to the next queue with a switchover rate of ε.

The following parameters are used:

N = number of queues in the system
λ_i = arrival rate of packets offered to queue i; $i = 1,\ldots,N$
S_i = capacity of queue i; $i = 1,\ldots,N$
μ = mean service rate of the server
ε = mean switchover rate of the server

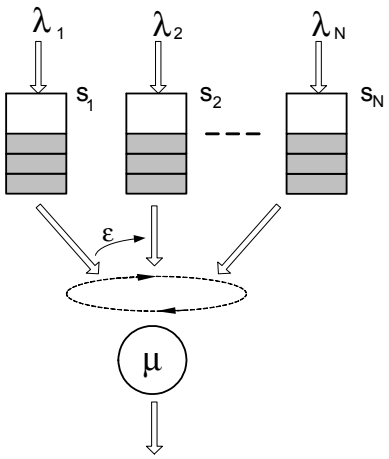


Fig. 4. System model for a cyclic service queueing system

4.2 Basic two-queue system

The analysis of cyclic service queueing systems is presented with a model that has only two queues as shown in Figure 5. Such a system can be considered as an $M/M/1$-s system.
The two-queue cyclic service system consists of one server and two queues with a capacity of s_1 and s_2 respectively, as shown in Figure 5. The mean arrival rates to the two queues are given by λ_1 and λ_2 respectively, while server completes each service with a mean rate of μ.

4.2.1 Analysis

For an exact analysis, the system states can be described by a vector $\{Q_1(t), Q_2(t), \ldots, Q_n(t), I(t), X(t)\}$, where $Q_i(t)$ is the number of packets in the ith queue, $I(t)$ is the current location of the server within the cycle and $X(t)$ is the age of the current service (Kuehn, 1979). In this study, the single-stage service process is taken to be a Markov process having a mean rate of μ. $X(t)$ can then be ignored due to the PASTA (Poisson Arrivals See Time Averages) property of the service process, which leaves us the vector $\{Q_1(t), Q_2(t), \ldots, Q_n(t), I(t)\}$ that accurately describes the system states. Hence for this two-queue system, three variables for each system state are required – one each for the number of occupied queue places – while another to show which queue's customer is currently undergoing service. Each state is then defined by the vector $\{Q_1(t), Q_2(t), I(t)\}$, where $Q_1(t)$ is the number of customers in the system coming through the first queue, $Q_2(t)$ is the number of customers in the system coming through the second queue and $I(t)$ is the current location of the server within the cycle. Clearly, $I(t)$ can have only two values where a value of 1 means that the server is serving a customer from queue 1 while 2 means that the server is serving a customer from queue 2. $Q_1(t)$ and $Q_2(t)$ can vary from zero to s_1 and s_2, respectively. The state diagram will hence be three-dimensional as shown in Figure 6, where transitions along the x-axis show arrivals of customers from queue 1 while transitions along the y-axis show arrivals of customers from queue 2. The z-axis shows the current location of the server within the cycle, with the front xy-plane showing the service of packets from queue 1 and the back xy-plane showing the service of packets from queue 2. State diagram of

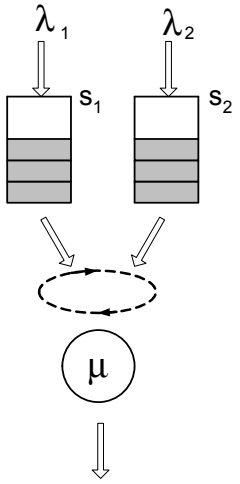


Fig. 5. System diagram for a two-queue cyclic service system

such systems usually consists of two parts – a boundary portion and a repeating portion. The boundary portion usually shows the states and transitions when the queues of the system are either empty or full, while the repeating portion usually shows the states and transitions when there is something in the queues but the queues are still not full. For very large state diagrams, such a depiction is very useful in studying the behavior of the system. Figure 7 shows a simplified view of the repeating portion of the state diagram in which transitions to and from just one state are shown. The server will switch from one queue to the other with a mean rate of ε.
Using the state diagram, the state probabilities p_i of all the states can be calculated by solving the system of linear equations. Using these state probabilities, the mean number in system and mean number in queue can then be found using the following equations.
Mean number of customers in system:

$$E[N] = \sum_{x=0}^{s+1} x p_x \tag{1}$$

Mean number of customers in queue:

$$E[Q] = \sum_{x=1}^{s+1} (x-1) p_x \tag{2}$$

From these equations, using the Little's theorem (Little, 1961), we get
Mean time in system:

$$T_S = \frac{E[N]}{\lambda} \tag{3}$$

Mean waiting time:

$$T_W = \frac{E[Q]}{\lambda} \tag{4}$$

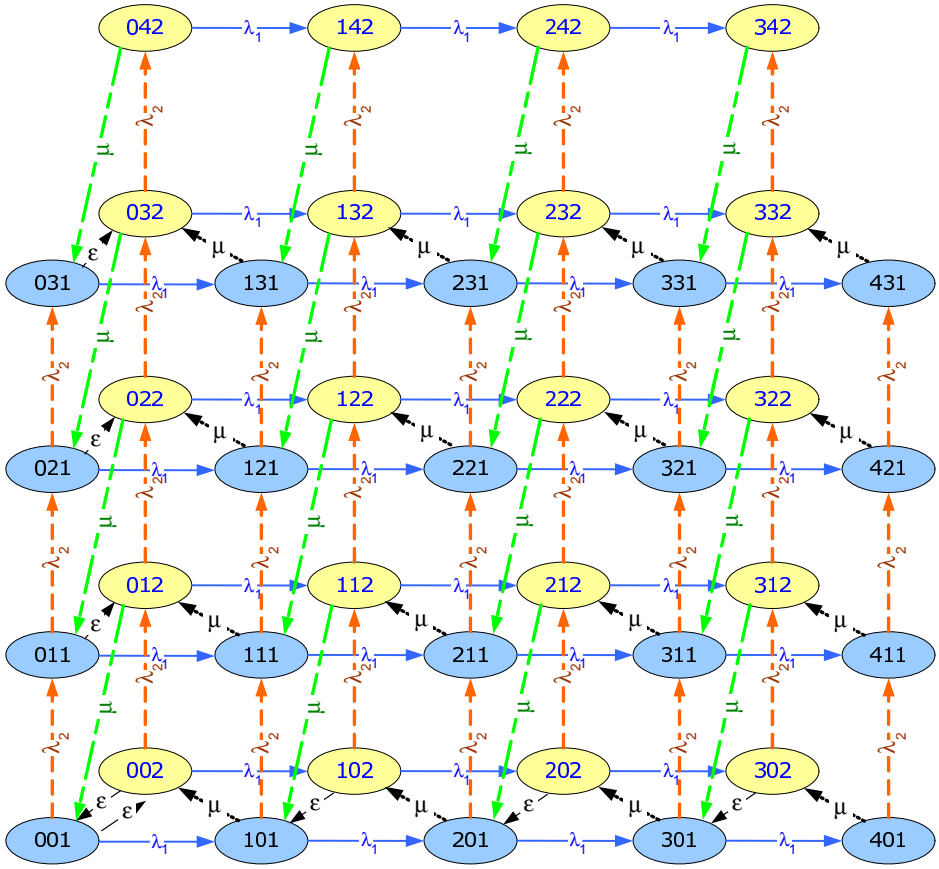


Fig. 6. State diagram for a two-queue cyclic service system

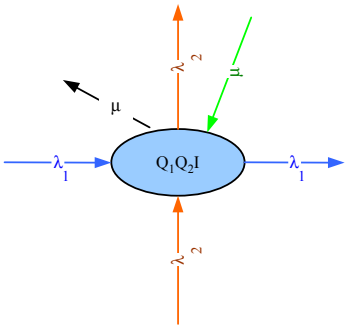


Fig. 7. Simplified view of the transitions to and from a state for a two-queue cyclic service system

An important point to note here is that the number in system are being considered, i.e., number in queue plus any customer that may be in service, and not just the number in queue. Hence in the state diagram of Figure 6 as well as the equations, Q_1 goes from 0 to $s_1 + 1$ and not s_1, while Q_2 goes from 0 to $s_2 + 1$ and not s_2.
Using (1) to (4), the various characteristic measures can be calculated for each queue as given in (5) to (8).

$$E[N_1] = \sum_{i_2=0}^{s_2} \sum_{i_1=0}^{s_1+1} i_1 P(i_1, i_2, 1) + \sum_{i_2=0}^{s_2+1} \sum_{i_1=0}^{s_1} i_1 P(i_1, i_2, 2) \tag{5}$$

$$E[Q_1] = \sum_{i_2=0}^{s_2} \sum_{i_1=2}^{s_1+1} (i_1 - 1) P(i_1, i_2, 1) + \sum_{i_2=0}^{s_2+1} \sum_{i_1=1}^{s_1} i_1 P(i_1, i_2, 2) \tag{6}$$

$$T_{S_1} = \frac{E[N_1]}{\lambda_1} \tag{7}$$

$$T_{W_1} = \frac{E[Q_1]}{\lambda_1} \tag{8}$$

When a customer arrives in a system and finds the server busy, it has to wait. If all the probabilities for the states in which the customer has to wait are summed up, the probability of waiting is obtained. Similarly, when a customer arrives to a system and finds the queue full, it will be blocked. If all the probabilities of such states are summed, the probability of blocking is obtained. The probabilities of waiting and blocking for this system are as follows:

$$W_1 = \sum_{i_2=0}^{s_2} \sum_{i_1=1}^{s_1} P(i_1, i_2, 1) + \sum_{i_2=0}^{s_2+1} \sum_{i_1=0}^{s_1-1} P(i_1, i_2, 2) \tag{9}$$

$$B_1 = \sum_{i_2=0}^{s_2} P(s_1 + 1, i_2, 1) + \sum_{i_2=0}^{s_2+1} P(s_1, i_2, 2) \tag{10}$$

4.2.2 Results

The various characteristic measures for customers in queue 1 will be affected not only by the queue length and arrival rate in queue 1, but also the arrival rate and maximum queue size of queue 2. Similarly, the switchover rate, although ignored during service, may still have an effect on the characteristic measures, especially at lower arrival rates and needs to be studied further.
In order to study these effects, various characteristic measures for customers in queue 1 given by (5) to (10) are plotted against arrival rate in queue 1 for different queue sizes and different arrival rates in queue 2. Symmetric as well as asymmetric traffic loads and queue sizes for both queues are studied.
Figure 8 shows the mean number of customers in queue 1 against varying arrival rate in queue 1, for various queue capacities. The graph shows that the mean number of customers in queue 1 increases slowly for low arrival rates up to 0.4, but increases rapidly from 0.4 to 0.7. It then stabilizes and levels out after the saturation point (arrival rate of 1.0). The graph also shows that increasing the capacity in queue 2 from 3 to 10 has a very small effect on the mean number of customers in queue 1. On the other hand, Figure 9 shows the mean number of customers

in queue 1 against varying arrival rate in queue 1, for various arrival rates in queue 2. It can be clearly seen that the arrival rate of queue 2 has a significant effect on the queue length distribution in queue 1. At low arrival rates in queue 2, the rate of increase in the queue length of queue 1 is much slower than the rate of increase observed for a high arrival rate in queue 2, as on average, the server spends more time serving customers of queue 2, especially at lower arrival rates of queue 1.

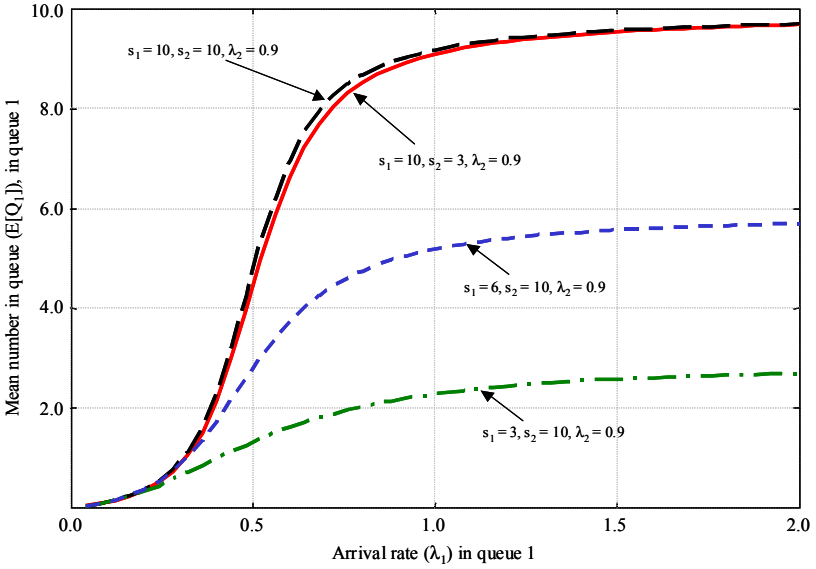


Fig. 8. Effect of varying queue sizes of queues 1 and 2 on number of customers in queue 1 for a two-queue system

Figures 10 and 11 show the mean waiting time for customers of queue 1 against the arrival rate of customers in queue 1, for varying queue capacities of both queues and varying arrival rate of customers in queue 2. Here again, a similar behavior is seen, whereby the queue capacity of queue 2 has a very small effect on the waiting time of customers in queue 1, as shown in Figure 10, but the increase of the arrival rate in queue 2 significantly increases the mean waiting time of customers in queue 1.

Finally, in Figures 12 and 13, the effect of queue 2 on the probability of blocking and the probability of waiting for customers in queue 1 is observed. Only the effect of increasing the arrival rate in queue 2 are shown as it has been observed that queue capacity of queue 2 has little effect on measures of queue 1. Here again, it is observed that a lower arrival rate in queue 2 results in a gradual increase in the blocking and waiting for customers of queue 1 as compared to a higher arrival rate, in which case this increase is quite abrupt.

4.3 Generalization to n-queue systems

An n-queue cyclic service system requires $n+1$ state variables to describe a state and hence, an $n+1$ dimensional state diagram. An important feature that is observed in these systems is the symmetry of the model. Extending the two-queue model to a more general n-queue model

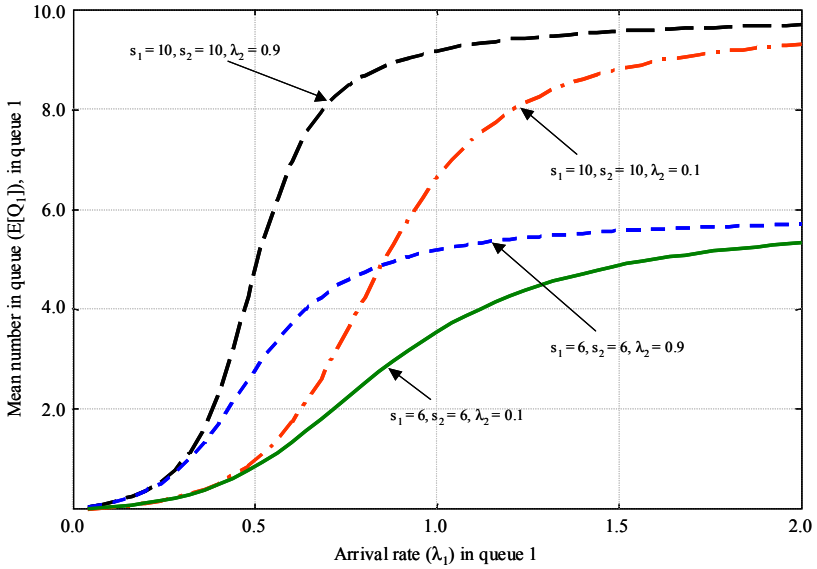


Fig. 9. Effect of varying arrival rate to queue 2, on number of customers in queue 1 for a two-queue system

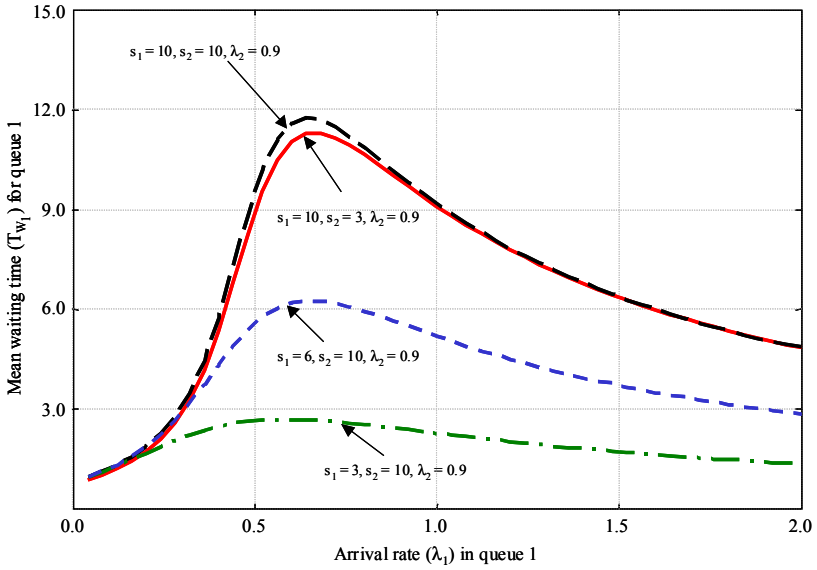


Fig. 10. Effect of varying queue sizes of queues 1 and 2, on waiting time of customers in queue 1 for a two-queue system

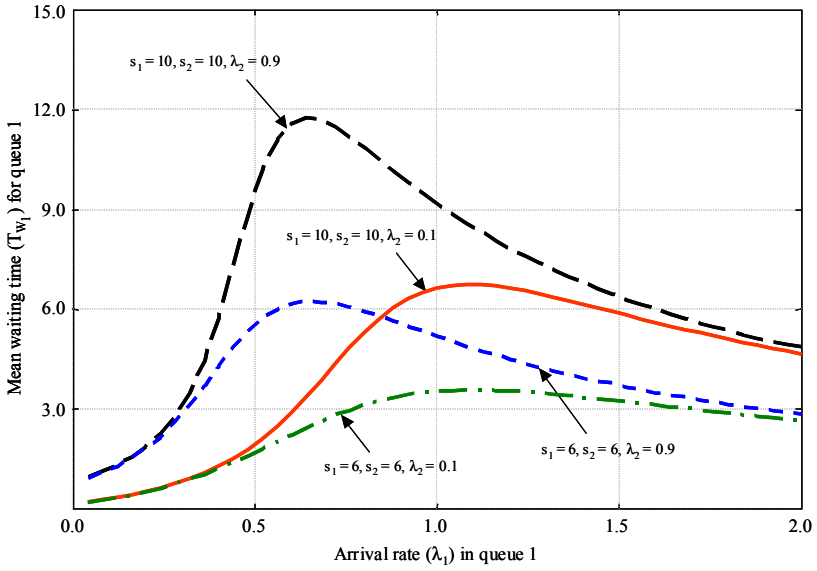


Fig. 11. Effect of varying arrival rate to queue 2, on waiting time of customers in queue 1 for a two-queue system

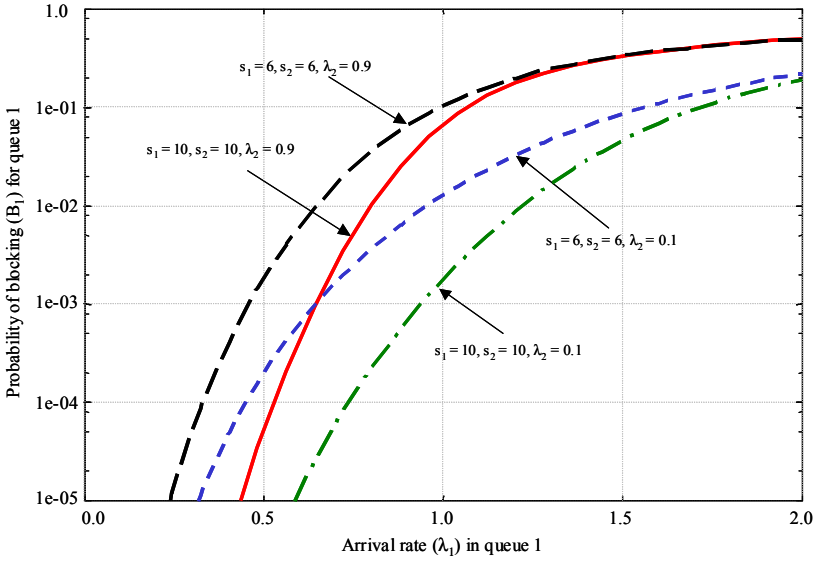


Fig. 12. Effect of varying arrival rate and maximum queue size of queue 2 on probability of blocking for customers in queue 1, for a two-queue system

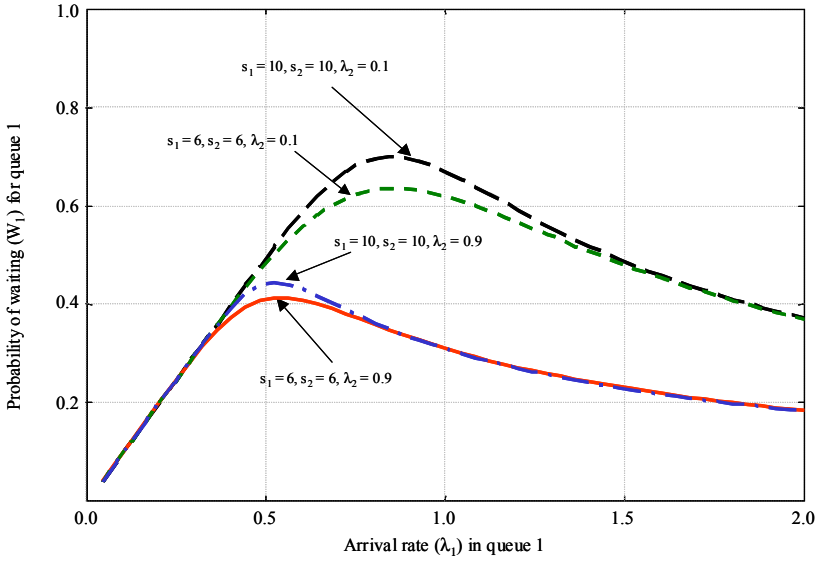


Fig. 13. Effect of varying arrival rate and maximum queue size of queue 2, on probability of waiting for customers in queue 1, for a two-queue system

is quite straight-forward. The complex part is the difficulty in drawing a state diagram with more than three dimensions. Due to the symmetry of the model, however, it is quite sufficient to draw a subset of the diagram for the boundary portion and the repeating portion of the system. The derivation of the system equations is also straightforward and (11) to (16) give the various measures for an n-queue system with switchover time ignored during service. The mean number in system and mean number in queue are given by:

$$\begin{aligned} E[N_1] = &\sum_{i_n=0}^{s_n} \cdots \sum_{i_2=0}^{s_2} \sum_{i_1=0}^{s_1+1} i_1 P(i_1, i_2, \cdots, i_n, 1) \\ &+ \sum_{i_n=0}^{s_n} \cdots \sum_{i_2=0}^{s_2+1} \sum_{i_1=0}^{s_1} i_1 P(i_1, i_2, \cdots, i_n, 2) \\ &+ \cdots + \sum_{i_n=0}^{s_n+1} \cdots \sum_{i_2=0}^{s_2} \sum_{i_1=0}^{s_1} i_1 P(i_1, i_2, \cdots, i_n, n) \end{aligned} \tag{11}$$

$$
\begin{aligned}
E[Q_1] = & \sum_{i_n=0}^{s_n} \cdots \sum_{i_2=0}^{s_2} \sum_{i_1=2}^{s_1+1} (i_1 - 1) P(i_1, i_2, \cdots, i_n, 1) \\
& + \sum_{i_n=0}^{s_n} \cdots \sum_{i_2=0}^{s_2+1} \sum_{i_1=1}^{s_1} i_1 P(i_1, i_2, \cdots, i_n, 2) \\
& + \cdots + \sum_{i_n=0}^{s_n+1} \cdots \sum_{i_2=0}^{s_2} \sum_{i_1=1}^{s_1} i_1 P(i_1, i_2, \cdots, i_n, n)
\end{aligned} \tag{12}
$$

Using Little's theorem, the mean time in system and the mean waiting time can be obtained as follows:

$$T_{S_1} = \frac{E[N_1]}{\lambda_1} \tag{13}$$

$$T_{W_1} = \frac{E[Q_1]}{\lambda_1} \tag{14}$$

The probability of waiting and probability of blocking can be calculated from the following equations.

$$
\begin{aligned}
W_1 = & \sum_{i_n=0}^{s_n} \cdots \sum_{i_2=0}^{s_2} \sum_{i_1=1}^{s_1} P(i_1, i_2, \cdots, i_n, 1) \\
& + \sum_{i_n=0}^{s_n} \cdots \sum_{i_2=0}^{s_2+1} \sum_{i_1=0}^{s_1-1} P(i_1, i_2, \cdots, i_n, 2) \\
& + \cdots + \sum_{i_n=0}^{s_n+1} \cdots \sum_{i_2=0}^{s_2} \sum_{i_1=0}^{s_1-1} P(i_1, i_2, \cdots, i_n, n)
\end{aligned} \tag{15}
$$

$$
\begin{aligned}
B_1 = & \sum_{i_n=0}^{s_n} \cdots \sum_{i_2=0}^{s_2} P(s_1 + 1, i_2, \cdots, i_n, 1) \\
& + \sum_{i_n=0}^{s_n} \cdots \sum_{i_2=0}^{s_2+1} P(s_1, i_2, \cdots, i_n, 2) \\
& + \cdots + \sum_{i_n=0}^{s_n+1} \cdots \sum_{i_2=0}^{s_2} P(s_1, i_2, \cdots, i_n, n)
\end{aligned} \tag{16}
$$

5. Systems with non-zero switchover times

Cyclic service queueing systems have a broad range of applications in communication systems mainly as a means of providing fairness to incoming traffic. In such systems, the server is required to switch to the next traffic stream after serving one. Usually in optical networks, this switching is done very fast as compared to the service and hence, the switchover time – which is the time taken by the server to switch from one stream to the next – is usually ignored. This however, can cause large differences in the results, especially if the switchover time is not very small as compared to the service time.
This section presents an analytical study of an edge node that employs non-exhaustive cyclic service to serve various incoming streams as a two-stage process (serving and switchover), and finite size queues to model real systems as closely as possible. The effect of switchover time on the performance of such systems is then studied. Comparison of systems with various ratios of switchover times to the service times is also done and the scenarios under which switchover times cannot be ignored are discussed.

5.1 Cyclic service system with two-stage service

An n-queue cyclic service system with two-stage non-exhaustive service can be modelled as shown in Figure 14. The n queues, each of size $s_i (i = 1, \ldots, n)$, are served in a round-robin manner by a server with a negative exponentially distributed service rate of mean μ. The arrival rate to each queue is Poisson with mean $\lambda_i (i = 1, \ldots, n)$. The average time taken by the server to switch over from one queue to the next is given by $1/\varepsilon$ where ε is the mean switchover rate.
At each scanning epoch, the server processes one packet in the queue if there is at least one packet waiting, with a rate of μ and then switches over to the next queue with a rate of ε. In case there is no waiting packet in the queue, the server simply switches over to the next queue with a switchover rate of ε.

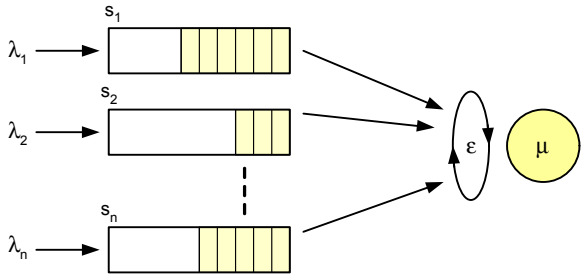


Fig. 14. System model for a queueing system with two-stage cyclic service

5.1.1 Analysis

For an exact analysis, the system states can be described by a vector $\{Q_1(t), Q_2(t), \ldots, Q_n(t), I(t), X(t)\}$, where $Q_i(t)$ is the number of packets in the ith queue, $I(t)$ is the current location of the server within the cycle and $X(t)$ is the age of the current service (Kuehn, 1979). In this study, a two-stage service process is assumed, with each of its two stages as a Markov process having mean rates of μ and ε for service and switchover,

respectively. $X(t)$ can then be ignored due to the PASTA (Poisson Arrivals See Time Averages) property of the service process, which leaves us the vector $\{Q_1(t), Q_2(t), \ldots, Q_n(t), I(t)\}$ that accurately describes the system states.
In case of the two-stage process presented here however, an additional state is needed to differentiate between two switchover cases. The first occurs after the processing of the packet in a non-empty queue, in which case the service will be two-stage; and the switchover that occurs for an empty queue, in which case the service will consist of only the switchover. The state space is hence modelled using the vector $\{Q_1(t), Q_2(t), \ldots, Q_n(t), I(t), K(t)\}$, where $K(t)$ is defined as the status of the server.
This study is restricted to two queues ($n = 2$). Each state is described by four parameters (i_1, i_2, j, k), where i_1 is the number of packets in queue 1, i_2 is the number of packets in queue 2, j is the current location of the server within the cycle ($j = 1$ means the server is scanning queue 1 while $j = 2$ means the server is scanning queue 2) and k is the status of the server. This server status can have three possible values: $k = 0$ means that the server encountered an empty queue and will simply switchover to the next queue without processing a packet, $k = 1$ means that the server is processing a packet, and $k = 2$ means that the server is switching over to the next queue after processing a packet in a non-empty queue. The resulting three-dimensional state diagram with a queue capacity of two for each queue is shown in Figure 15. It can be easily seen that a state diagram for more than two queues needs a four-dimension representation, which is not possible to draw on paper.
Figure 16 shows a simplified view of the repeating portion of the state diagram in which transitions to and from just one state are shown. Note that these transitions can be divided into four main parts:

- Arrivals to queue 1, which result in an increment of i_1. All other parameters describing the state remain unchanged.
- Arrivals to queue 2, which result in an increment of i_2. All other parameters remain unchanged.
- Service followed by switchover from queue 1. In this case, processing of a packet results in a decrement of i_1, while the server status changes to 2 from 1. Switchover then changes j to 2 from 1, indicating that server is now pointing at queue 2, while the server status k changes to 1 from 2 indicating that the server is ready to serve queue 2.
- Service followed by switchover from queue 2. In this case, processing of a packet results in a decrement of i_2, while the server status k changes to 2 from 1. Switchover then changes j to 1 from 2, indicating that the server is now pointing at queue 1, while the server status changes to 1 from 2 indicating that the server is ready for service.

Figure 17 shows the states of the system and the transitions that occur when the queues are empty. Note that in this case, no packets will be processed and the service process will consist of just the switchover. The transitions would be similar to those described above with the exception that after service and switchover, if the server sees an empty queue, the server status k will go to 0 instead of 1.
The steady state solution of the state-space model shown in Figure 15 can be obtained by solving the system of equations that is obtained from the state diagram. The steady state probabilities thus obtained can be used to solve for the various system measures. The mean number in the system, $E[N_1]$ and the mean number in queue $E[Q_1]$ for packets of type 1, i.e., packets that arrive to queue 1 are given by (17) and (18), respectively.

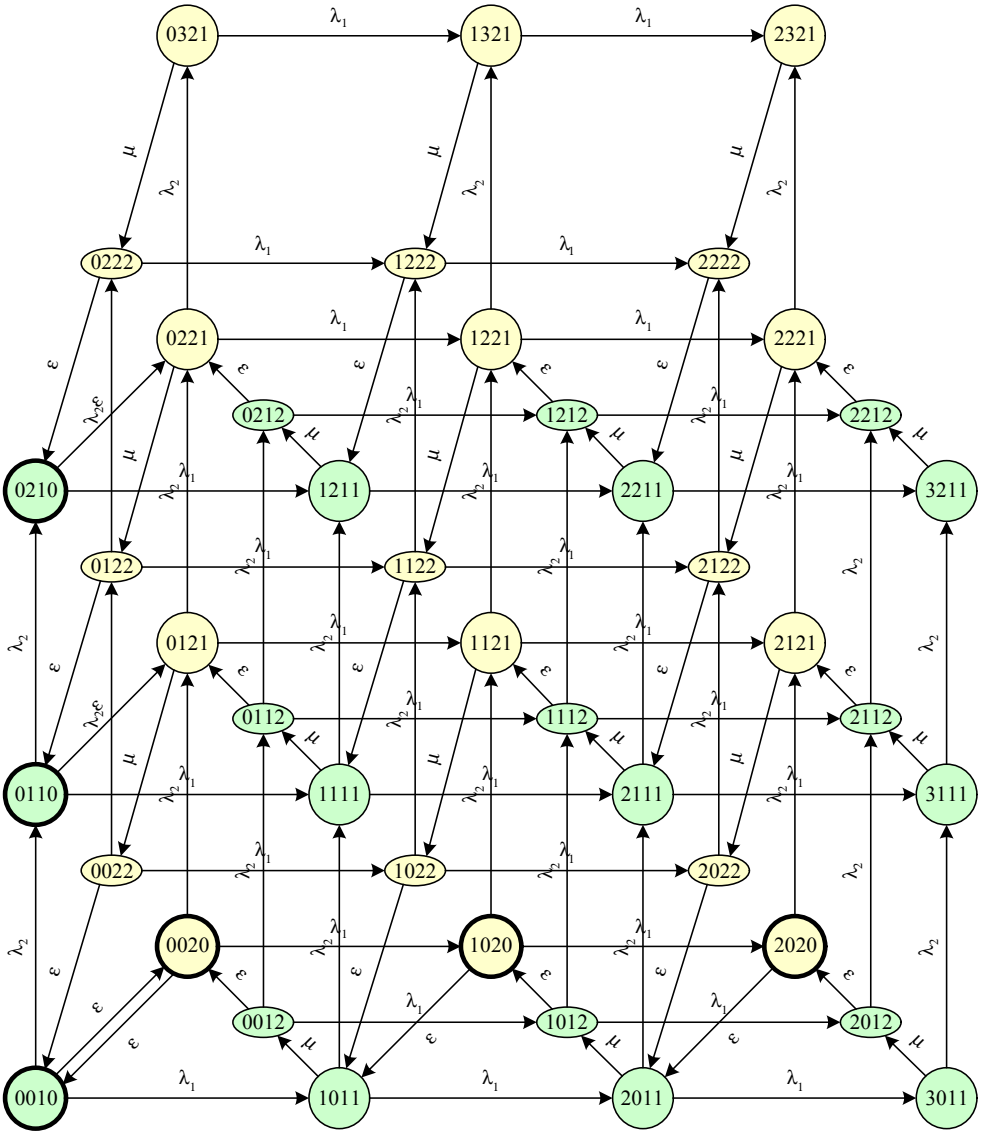


Fig. 15. Three dimensional state diagram of a non-exhaustive cyclic queueing system with two queues, each of size two and a two-stage service process

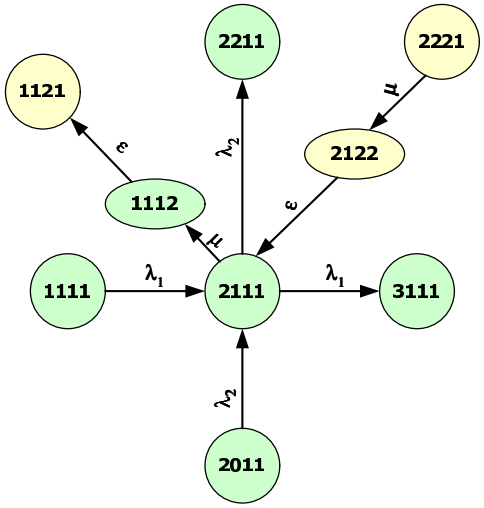


Fig. 16. Simplified view of the transitions to and from a state when queue are not empty

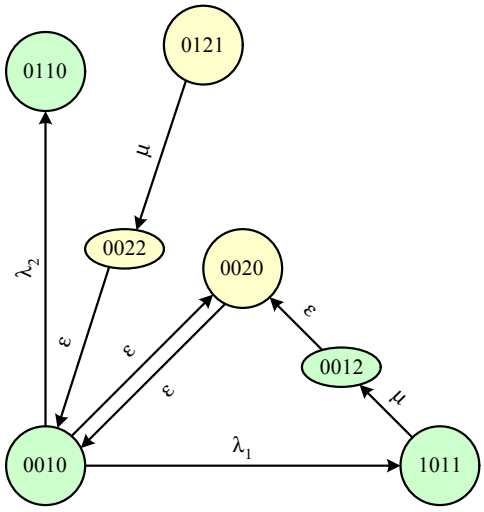


Fig. 17. Simplified view of the transitions to and from a state when queue are empty

$$E[N_1] = \sum_{i_2=0}^{s}\sum_{i_1=1}^{s+1} i_1 P(i_1, i_2, 1, 1) + \sum_{i_1=0}^{s}\left(i_1 P(i_1, 0, 2, 0) + \sum_{i_2=1}^{s+1} i_1 P(i_1, i_2, 2, 1)\right) + \sum_{j=1}^{2}\sum_{i_2=0}^{s}\sum_{i_1=0}^{s} i_1 P(i_1, i_2, j, 2) \tag{17}$$

$$E[Q_1] = \sum_{i_2=0}^{s}\sum_{i_1=1}^{s+1} (i_1 - 1) P(i_1, i_2, 1, 1) + \sum_{i_1=0}^{s}\left(i_1 P(i_1, 0, 2, 0) + \sum_{i_2=1}^{s+1} i_1 P(i_1, i_2, 2, 1)\right) + \sum_{j=1}^{2}\sum_{i_2=0}^{s}\sum_{i_1=0}^{s} i_1 P(i_1, i_2, j, 2) \tag{18}$$

Applying Little's law (Little, 1961), the mean system time, T_{S1}, and mean waiting time, T_{W1}, can be found from (17) and (18) as follows:

$$T_{S_1} = \frac{E[N_1]}{\lambda_1} \tag{19}$$

$$T_{W_1} = \frac{E[Q_1]}{\lambda_1} \tag{20}$$

The probability of waiting, W_1, is obtained by summing up the probabilities of the state, where on arrival, a packet has to wait, and is given by (21) for the packets of queue 1, while the probability of blocking, B_1, is obtained by summing up the state probabilities where queue 1 is full as given in (22).

$$W_1 = \sum_{i_2=0}^{s}\sum_{i_1=1}^{s} P(i_1, i_2, 1, 1) + \sum_{i_1=0}^{s-1}\left(P(i_1, 0, 2, 0) + \sum_{i_2=1}^{s+1} P(i_1, i_2, 2, 1)\right) + \sum_{j=1}^{2}\sum_{i_2=0}^{s}\sum_{i_1=0}^{s-1} P(i_1, i_2, j, 2) \tag{21}$$

$$B_1 = \sum_{i_2=0}^{s} P(s+1, i_2, 1, 1) + P(s, 0, 2, 0) + \sum_{i_2=1}^{s+1} P(s, i_2, 2, 1) + \sum_{j=1}^{2}\sum_{i_2=0}^{s} P(s, i_2, j, 2) \tag{22}$$

5.1.2 Results

The effect of varying the switchover rate and input load on the mean number in system, mean waiting time, probability of waiting and probability of blocking is studied. In Figure 18, the input load for queue 2 is fixed at 0.1 and queue size of both queues to 10. The resulting graph shows that for values of switchover rate comparable to the service rate, the queue capacity is reached quickly and at a much lower load as compared to when the switchover rate is ten times that of the service rate. This effect is significantly reduced when the switchover rate is increased to hundred times that of the service rate and beyond. The same effect can also be noted in Figure 19 that shows the mean waiting time against the arrival rate for queue 1.
In Figures 20 and 21, the load of queue 2 is also varied from 0.1 to 0.9 along with the switchover rate to service rate ratio from 1 to 10, to see their combined effect. An interesting phenomenon to note here is that when the switchover rate is equal to the service rate, the effect of varying the load of queue 2 does not significantly affect the mean number in queue or the mean waiting time for queue 1. However, when the switchover rate is ten times faster than the service rate, the effect of varying the load in queue 2 has a significant impact on the mean number in queue and the mean waiting time for customers in queue 1. Note that the mean waiting time increases rapidly with increasing traffic until a certain level, after which the overload in the system results in blocking, thus reducing the overall waiting.
Figure 22 shows the probability of waiting for customers in queue 1 for an arrival rate in queue 2 of 0.1. Probability of waiting is the probability that a customer, on entering the system, finds the server busy and has to wait in queue. Again it is observed that the effect of switchover rate is dominant when it is equal to the service rate but its impact is reduced as it is increased to 10 or 100 times the service rate.
Finally, Figure 23 shows the probability of blocking against the arrival rate of queue 1. The probability of blocking is the probability that the customer, on entering the system, finds the server and all queues full and is lost. Here, the switchover to service rate ratio as well as the arrival rate of queue 2 is varied again and it is observed that varying the arrival rate has little effect on the probability of blocking when the switchover and service rates are the same. However, at higher ratios of the switchover rate to service rate, the effect of varying the load on queue 2 has a big impact on the probability of blocking for customers in queue 1.
It can be concluded that switchover time should not be ignored in systems where the ratio of service time to switchover time is small (or conversely, the ratio of switchover rate to service rate is small) as it significantly affects the performance of the system. It is also observed that at switchover rates comparable to those of the service rates, the effect of varying arrival rates in the other queues has little effect on the system performance, but this effect becomes more pronounced as the ratio between switchover rate and service rate increases. Hence for high speed optical communication systems, like edge nodes that map Ethernet over SDH/SONET or burst assemblers in optical burst switching nodes, one should proceed with care whenever switchover times are involved as the high data rates usually mean that the ratio between the switchover rate and service rate might not be high enough to ignore the switchover rate during analysis.

6. Comparing systems with and without switchover times

In the previous section, the effect of the increase and decrease in the switchover time on the various characteristic measures was studied. This sections compares identical cyclic service queueing systems with and without switchover times. Hence the results from Sections 5.4.2 and 5.5.1 will be reused to make this comparison.

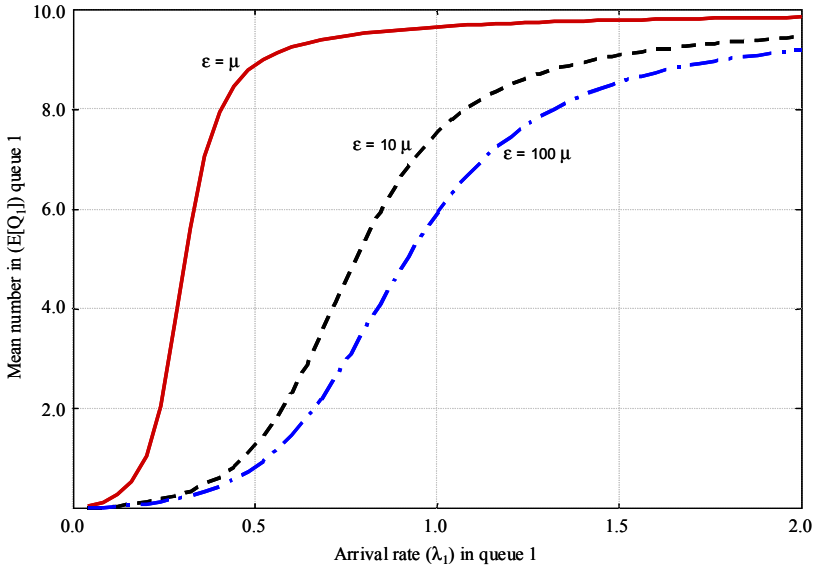


Fig. 18. Effect of varying the switchover rate on number of customers in queue 1

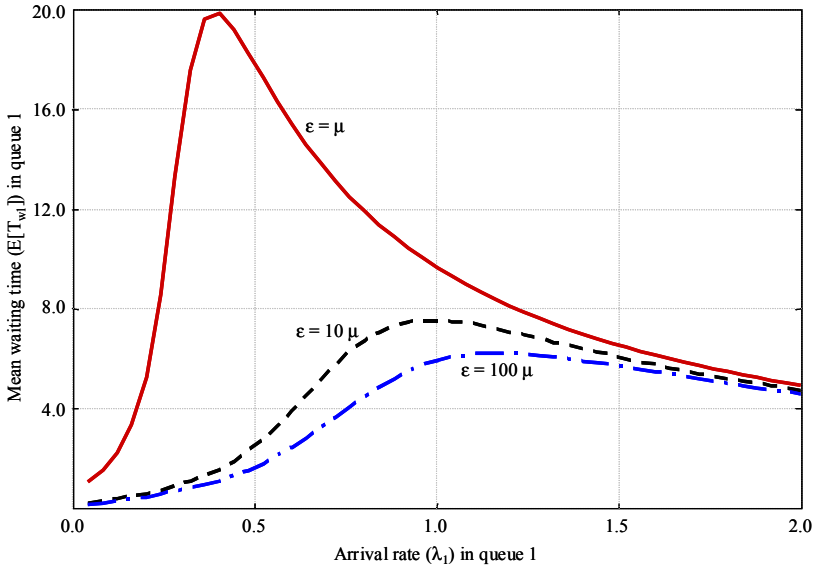


Fig. 19. Effect of varying the switchover rate on mean waiting time for customers in queue 1

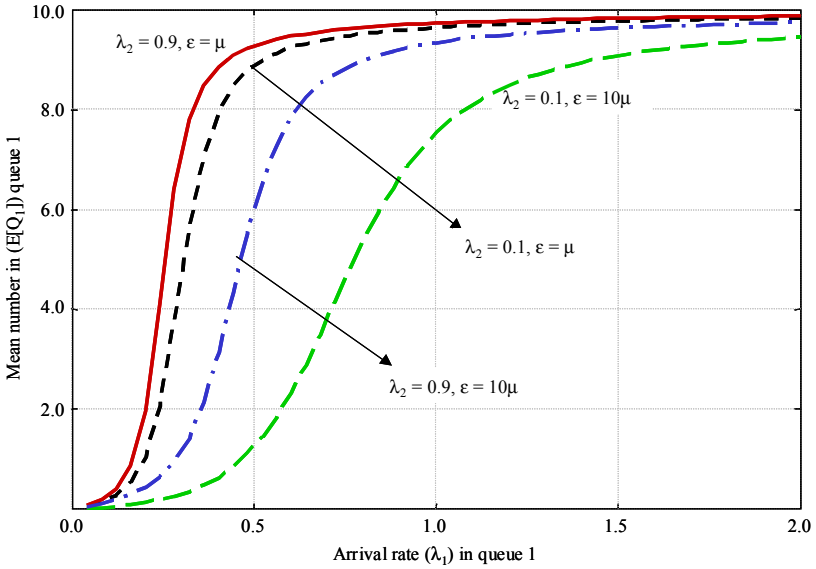


Fig. 20. Effect of varying the switchover rate and the arrival rate for queue 2 on number of customers in queue 1

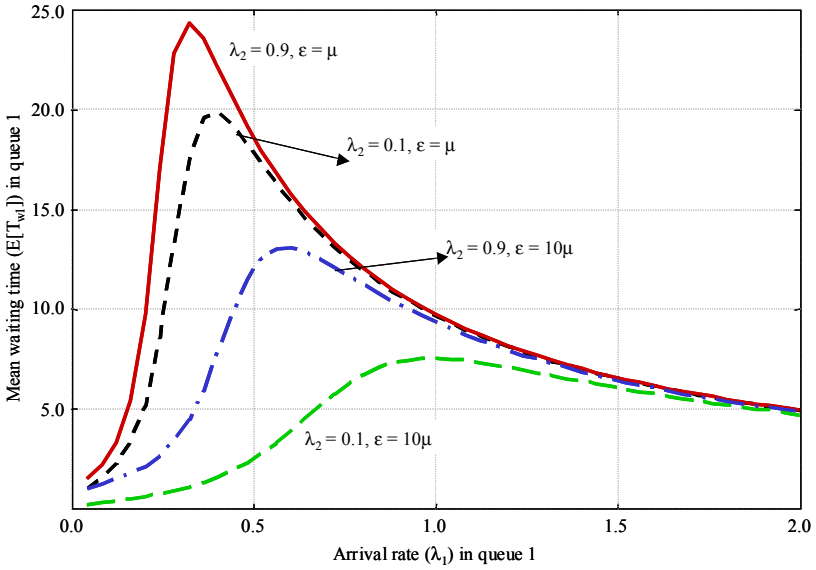


Fig. 21. Effect of varying the switchover rate and the arrival rate for queue 2 on mean waiting time for customers in queue 1

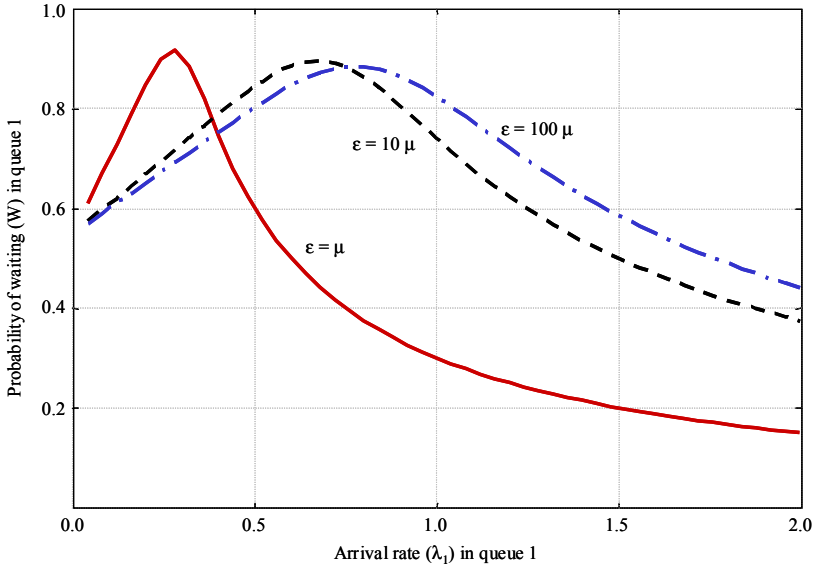


Fig. 22. Effect of varying the switchover rate on probability of waiting for customers in queue 1

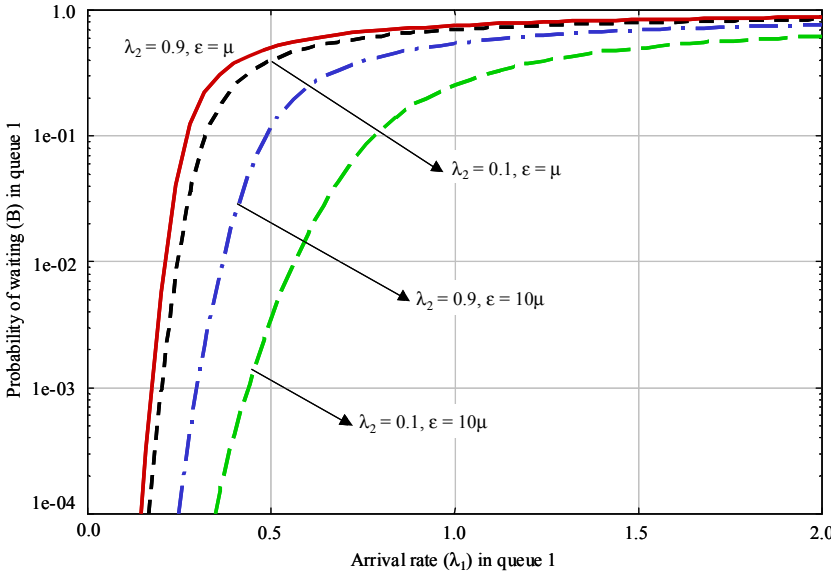


Fig. 23. Effect of varying the switchover rate and the arrival rate for queue 2 on probability of blocking for customers in queue 1

Figure 24 shows two sets of plots for the mean number in queue for customers of queue 1. The first set of plots is for systems in which the switchover rate is ignored during service. The second set of plots is for systems in which this rate is not ignored. These two sets of plots are drawn for switchover rates of 1, 10 and 100, respectively. It is clearly observed that for a switchover rate of 100 times the service rate, the plots for these two cases are almost identical. The difference, however, is not negligible when the switchover rate is decreased to 10 times the service rate. This difference becomes very significant when the switchover rate is of the order of the service rate. This shows that although for higher ratios of the switchover rate versus the service rate, it is safe to ignore the switchover rate during service, however, as this ratio decreases, the difference in values of the characteristic measures becomes too significant for the switchover time to be ignored.

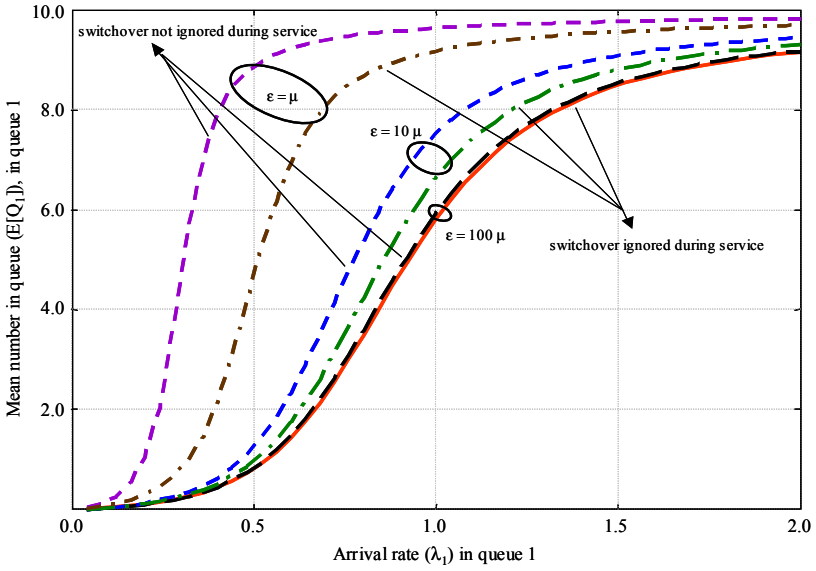


Fig. 24. Comparison of the mean number in queue 1 for systems with and without switchover rates during service, for a queue size of 10 and arrival rate of 0.1 in queue 2

The same phenomenon can be observed in Figure 25, which shows two sets of plots for the mean waiting time for customers of queue 1. Here again, the first set of plots is for systems in which the switchover rate is ignored during service while the second set of plots is for systems in which this rate is not ignored. These two sets of plots are drawn for switchover rates of 1, 10 and 100, respectively. Again, it can be observed that for higher ratios of the switchover rate versus the service rate, it is safe to ignore the switchover rate during service, however, as this ratio decreases, the difference in values of the characteristic measures becomes too significant for the switchover time to be ignored. Typically for these systems, ratios of the switchover rate versus the service rate of more than 100 should be sufficiently large for the switchover time to be ignored. If this ratio is less than 100, ignoring the switchover time could lead to large differences in results.

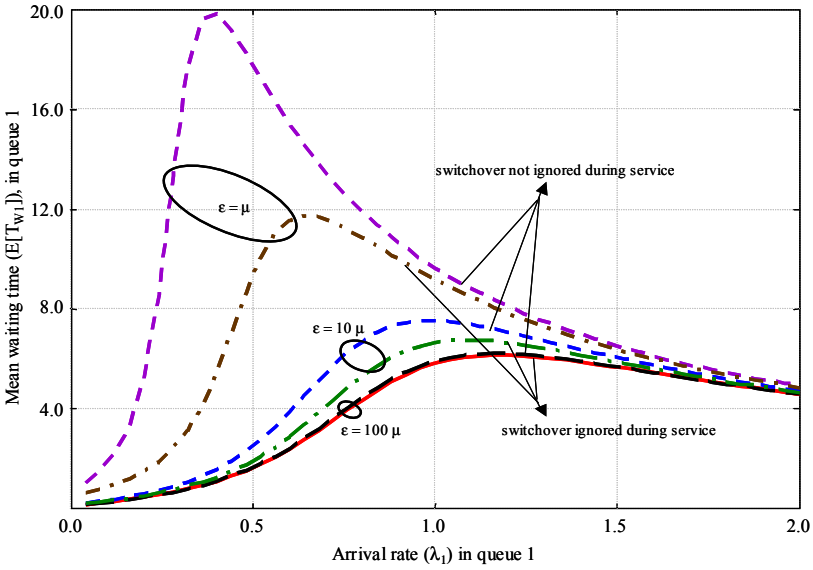


Fig. 25. Comparison of the mean waiting time in queue 1 for systems with and without switchover rates during service, for a queue size of 10 and arrival rate of 0.1 in queue 2

7. Summary

This chapter presented an overview of the various types of polling systems. Polling systems were classified and the existing work was summarized. Cyclic service queueing systems and their applications in modern day communication systems were then discussed. While a lot of work has been done on polling systems with exhaustive service and infinite queues, with several closed form solutions, the work on finite queue, non-exhaustive cyclic polling systems is very limited, and only approximate solutions are available. Starting with a simple two-queue cyclic polling model with switchover time ignored during service, various characteristic measures were studied, including the mean waiting time and the blocking probability for the customers in the system. This simple two-queue model was then extended to an n-queue model and generalized formulae were developed.. In most of the studies, the switchover time – an important parameter – has been ignored. In order to see the effect of the switchover time, especially in optical communication systems where ever increasing speeds imply an ever diminishing ratio of service time to switchover time, a two-stage service model was developed for a two-queue system with service followed by switchover. This model was then compared with the model in which switchover time was ignored during service. Significant differences were noted when the ratio of service time to switchover time was small. However, this difference was negligible where the ratio between service time and switchover time was greater than 100. It can thus be concluded that it is not always safe to ignore the switchover times. It is important to note that the various techniques discussed here have been mostly for small systems with two, or three-queues. It is straightforward to extend this study to multiple queues with large queue sizes because of the symmetric nature of the systems. The practical limita-

tion is due to the state space explosion that occurs when large systems are modelled, which result in large computational times and require heavy computational resources.

8. References

Borst, S. C. & Boxma, O. J. (1997). Polling Models With and Without Switchover Times, *Operations Research* **45**(4): 536–543.

Boxma, O. J. (1989). Workloads and Waiting Times in Single-server Systems with Multiple Customer Classes, *Queueing Systems* **5**: 185–214.

Boxma, O. J. (2002). Two-Queue Polling Models with a Patient Server, *Operations Research* **112**: 101–121.

Bruneel, H. & Kim, B. G. (1993). *Discrete-time Models for Communication Systems Including ATM*, Boston: Kluwer.

Bux, W. & Truong, H. L. (1983). Mean-delay Approximation for Cyclic-Service Queueing Systems, *Performance Evaluation* **3**(3): 187–196.

Choi, S. (2004). Cyclic Polling Based Dynamic Bandwidth Allocation for Differentiated Classes of Service in Ethernet Passive Optical Networks, *Photonic Network Communications* **7**(1): 87–96.

Chung, H. U. C. & Jung, W. (1994). Performance Analysis of Markovian Polling Systems with Single Buffers, *Performance Evaluation* **19**(4): 303–315.

Coooper, R. B. & Murray, G. (1969). Queues Served in Cyclic Order, *The Bell Systems Technical Journal* **48**: 675–689.

Cooper, R. B. (1970). Queues Served in Cyclic Order : Waiting Times, *The Bell Systems Technical Journal* **49**: 399–413.

Cooper, R. B., Niu, S.-C. & Srinivasan, M. M. (1996). A Decomposition Theorem for Polling Models: The Switch-over Times Are Effectively Additive, *Operations Research* **44**(4): 629–633.

Eisenberg, M. (1971). Two Queues with Changeover Times, *Operations Research* **19**: 386–401.

Eisenberg, M. (1972). Queues with Periodic Service and Changeover Times, *Operations Research* **20**: 440/451.

Fuhrmann, S. W. (1992). A Decomposition Result for a Class of Polling Models, *Queueing Systems* **11**: 109/120.

Grillo, D. (1990). Polling Mechanism Models in Communication Systems - Some Application Examples, *Stochastic Analysis of Computer and Communication Systems, Amsterdam: North-Holland* pp. 659–698.

Hashida, O. (1972). Analysis of Multiqueue, *Review of the Electrical Communication Laboratories, Nippon Telegraph and Telephone Public Corporation* **20**(3): 189–199.

Ibe, O. C. & Trivedi, K. S. (1990). Stochastic Petri Net Models of Polling Systems, *IEEE Journal for Selected Areas in Communications* **8**(9): 1649–1657.

Jung, W. Y. & Un, C. K. (1994). Analysis of a Finite-buffer Polling System with Exhaustive Service Based on Virtual Buffering, *IEEE Transactions on Communications* **42**(12): 3144–3149.

Kuehn, P. J. (1979). Multiqueue Systems with Nonexhaustive Cyclic Service, *The Bell Systems Technical Journal* **58**(3): 671–699.

Lee, D.-S. (1996). A two-queue model with exhaustive and limited service disciplines, *Stochastic Models* **12**(2): 285–305.

Leibowitz, M. A. (1961). An Approximate Method for Treating a Class of Multiqueue Problems, *IBM J. Res. Develop.* **5**: 204–209.

Levy, H. & Sidi, M. (1990). Polling Systems: Applications, Modeling and Optimization, *IEEE Transactions on Communications* **38**(10): 1750–1760.

Little, J. D. C. (1961). A proof of the queueing formula L=λW, *Operations Research* **9**: 383–387.

Mack, C., Murphy, T. & Webb, N. (1957). The efficiency of N machines unidirectionally patrolled by one operative when walking times and repair times are constants, *J. Roy. Statist. Soc.* **B 19**: 166–172.

Magalhaes, M. N., McNickle, D. C. & Salles, M. C. B. (1998). Outputs from a Loss System with Two Stations and a Smart (Cyclic) Server, *Investigacion Oper.* **16**(1-3): 111–126.

Miorandi, D., Zanella, A. & Pierobon, G. (2004). Performance Evaluation of Bluetooth Polling Schemes: An Analytical Approach, *ACM Mobile Networks Applications* **9**(2): 63–72.

Nagle, J. B. (1987). On Packet Switches with Infinite Storage, *IEEE Transactions on Communications* **35**(4): 435–438.

Onvural, R. O. & Perros, H. G. (1989). Approximate Throughput Analysis of Cyclic Queueing Networks with Finite Buffers, *IEEE Transactions on Software Engineering* **15**(6): 800–808.

Srinivasan, M. M., Niu, S.-C. & Cooper, R. B. (1995). Relating Polling Models with Nonzero and Zero Switchover Times, *Queueing Systems* **19**: 149–168.

Takagi, H. (1986). *Analysis of Polling Systems*, The MIT Press, Cambridge, MA, chapter 2.

Takagi, H. (1988). Queuing analysis of polling models, *ACM Computing Surveys* **20**(1): 5–28.

Takagi, H. (1990). Queueing analysis of polling models: an update, *Stochastic Analysis of Computer and Communication Systems, Elsevier Science Publishers B. V. (North-Holland), Amsterdam* pp. 267–318.

Takagi, H. (1992). Analysis of an M/G/1//N Queue with Multiple Server Vacations, and its Application to a Polling Model, *J. Oper. Res. Soc. Japan* **35**: 300–315.

Takagi, H. (1997). Queueing analysis of polling models: progress in 1990-1994, *Frontiers in Queueing: Models and Applications in Science and Technology, CRC Press, Boca Raton, Florida* (Chapter 5): 119–146.

Takagi, H. (2000). Analysis and Applications of Polling Models, *Performance Evaluation* **LNCS-1769**: 423–442.

Takine, T., Takahashi, Y. & Hasegawa, T. (1986). Performance Analysis of a Polling System with Single Buffers and its Application to Interconnected Networks, *IEEE Journal on Selected Areas in Communication* **SAC-4**(6): 802–812.

Takine, T., Takahashi, Y. & Hasegawa, T. (1987). Analysis of a Buffer Relaxation Polling System with Single Buffers, *Proceedings of the Seminar on Queuing Theory and its Applications, May 11-13, Kyoto Univ., Kyoto, Japan* pp. 117–132.

Takine, T., Takahashi, Y. & Hasegawa, T. (1990). Modelling and Analysis of a Single-buffer Polling System Interconnected with External Networks, *INFOR.* **28**(3): 166–177.

Titenko, I. M. (1984). On Cyclically Served Multi-Channel Systems with Losses, *Avtom. Telemekh.* (10): 88–95.

Tran-Gia, P. (1992). Analysis of Polling Systems with General Input Process and Finite Capacity, *IEEE Transactions on Communications* **40**(2): 337–344.

Tran-Gia, P. & Raith, T. (1985a). Approximation for Finite Capacity Multiqueue Systems, *Conf. on Measurement, Modelling and Evaluation of Computer Systems, Dortmund, Germany* .

Tran-Gia, P. & Raith, T. (1985b). Multiqueue Systems with Finite Capacity and Nonexhaustive Cyclic Service, *International Seminar on Computer Networking and Performance Evaluation, Tokyo, Japan* .

Vishnevskii, V. M. & Semenova, O. V. (2006). Mathematical Models to Study the Polling Systems, *Automation and Remote Control* **67**(2): 173–220.

Vishnevsky, V. M., Lyakhov, A. I. & Bakanov, A. S. (1999). Method for Performance Evaluation of Wireless Networks with Centralized Control, *Proceedings of Distributed Computer Communication Networks (Theory and Applications), Tel-Aviv, Israel* pp. 189–194.

Vishnevsky, V. M., Lyakhov, A. I. & Guzakov, N. N. (2004). An Adaptive Polling Strategy for IEEE 802.11 PCF, *Proceedings of 7th International Symposium on Wireless Personal Multimedia Communications, Abano Terme, Italy* **1**: 87–91.

Ziouva, E. & Antonakopoulos, T. (2002/2007). Efficient Voice Communications over IEEE802.11 WLANs Using Improved PCF Procedures, *Proc. INC, Plymouth* .

Ziouva, E. & Antonakopoulos, T. (2003). Improved IEEE 802.11 PCF Performance Using Silence Detection and Cyclic Shift on Stations Polling, *IEE Proceedings Communications* **150**(1): 45–51.

7

Packet Dispatching Schemes for Three-Stage Buffered Clos-Network Switches

Janusz Kleban
Poznan University of Technology
Poland

1. Introduction

The continued growth of Internet Protocol-based traffic like data, voice, audio, TV, and gaming traffic, requires much more robust, highly scalable core routers/switches to handle the expected annual doubling of bandwidth in the United States and Europe and the expected tripling and possibly quadrupling of bandwidth in Asia. In the near future service providers will need to deploy a new class of core routers that have taken a major leap forward in design. While the bandwidth of external connections on core routers has increased in recent years from STM-1 to STM-16 and STM-64, tomorrow's core routers will need to support STM-256 connections operating at 40 Gbps. In addition, the number of line cards that the core router will need to support will grow dramatically to handle the aggregate subscriber and backbone bandwidth. To meet these new demands, tomorrow's router architectures will have to function very differently from those of today. They will require distributed memories and multi-stage switching fabrics that replace single-stage crossbars, allowing extraordinary scalability.
The main part of each high-performance network node is a switching fabric - instead of a shared central bus - which transfers a packet from its input link to its output link (Fig. 1). The switching fabric provides very fast transmission between line cards, therefore the router throughput is improved. Internally, high capacity switches/routers operate on fixed-size data units, called cells from the ATM jargon. This means that in the case of variable-size packets on transmission lines, as it is normally the case in the Internet, packets must be segmented into cells at switch inputs, and cells must be reassembled into packets at switch outputs (Chao & Cheuk, 2001). There are mainly two approaches to the implementation of high-speed packet switching systems. One approach is the single-stage switch architecture such as the crossbar switch, the other one is the multiple-stage switch architecture, such as the Clos-network switch. Most high-speed packet switching systems in the backbone of the Internet are currently built on the basis of a single-stage switching fabric with a centralized scheduler. Crossbar switches are internally nonblocking and simple in architecture. However, they are only little scalable due to the number of the crosspoints, which grows as N^2, where N is the total number of inputs/outputs. Multiple-stage Clos-network switches are a potential solution to overcome the limited scalability of single-stage switches, in terms of the number of input/output chip pins and the number of switching elements.

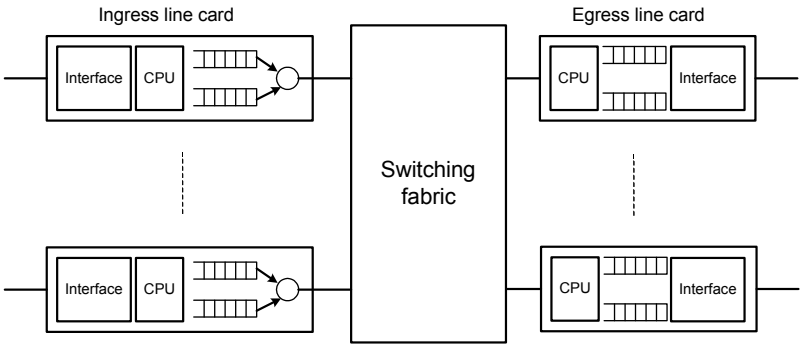


Fig. 1. High-performance router architecture

In the Clos-network switch packet scheduling is needed as there is a large number of shared resources where contention may occur. A cell transmitted within the multiple-stage Clos switching fabric can face internal blocking or output port contention. Internal blocking occurs when two or more cells contend for an internal link at the same time (Fig.2). A switch suffering from internal blocking is called blocking contrary to a switch that does not suffer from internal blocking called nonblocking. The output port contention occurs if there are multiple cells contend for the same output port.

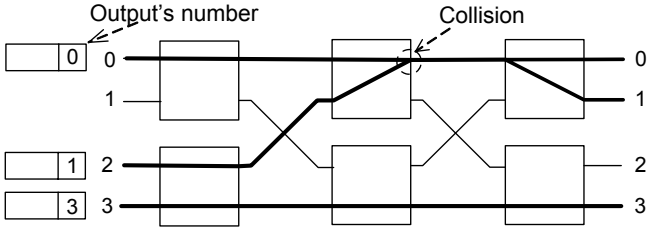


Fig. 2. Internal blocking: two cells destined for output ports 0 and 1 try to go through the same internal link, at the same time

Cells that have lost contention must be either discarded or buffered. Generally speaking, buffers may be placed at inputs, outputs, inputs and outputs, and/or within the switching fabric. Depending on the buffer placement respective switches are called input queued (IQ), output queued (OQ), combined input and output queued (CIOQ) and combined input and crosspoint queued (CICQ) (Yoshigoe &Christensen, 2003).

In the OQ strategy all incoming cells (i.e. fixed-length packets) are allowed to arrive at the output port and are stored in queues located at each output of switching elements. The cells destined for the same output port simultaneously do not face a contention problem because they are queued in the buffer at the output. To avoid the cell loss the system must be able to write N cells in the queue during one cell time. No arbiter is required because all the cells can be switched to respective output queue. The cells in the output queue are served using FIFO discipline to maintain the integrity of the cell sequence. In OQ switches the best performance (100% throughput, low mean time delay) is achieved, but every output port must be able to accept a cell from every input port simultaneously or at least within a single

time slot (a time slot is the duration of a cell). An output buffered switch can be more complex than an input buffered switch because the switching fabric and output buffers must effectively operate at a much higher speed than that of each port to reduce the probability of cell loss. The bandwidth required inside the switching fabric is proportional to both the number of ports *N* and the line rate. The internal speedup factor is inherent to pure output buffering, and is the main reason of difficulties in implementing switches with output buffering. Since the output buffer needs to store *N* cells in each time slot, its speed limits the switch size.

The IQ packet switches have the internal operation speed equal to (or slightly higher) than the input/output line speed, but the throughput is limited to 58,6% under uniform traffic and Bernoulli packet arrivals because of Head-Of-Line (HOL) blocking phenomenon (Chao & Cheuk, 2001). HOL blocking causes the idle output to remain idle even if at an idle input there is a cell waiting to be sent to an (idle) output. Due to other cell that is ahead of it in the buffer the cell cannot be transmitted over the switching fabric. An example of HOL blocking is shown in Fig. 3. This problem can be solved by selecting queued cells other than the HOL cell for transmission, but it is difficult to implement such queuing discipline in hardware. Another solution is to use speedup, i.e. the switch's internal links speed is greater than inputs/outputs speed. However, this also requires a buffer memory speed faster than a link speed. To increase the throughput of IQ switches space parallelism is also used in the switch fabric, i.e. more than one input port of the switch can transmit simultaneously.

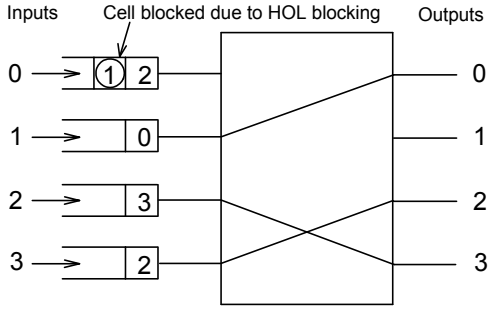


Fig. 3. Head-of-line blocking

The virtual output queuing (VOQ) is widely implemented as a good solution for input queued (IQ) switches, to avoid the HOL blocking encountered in the pure input-buffered switches. In VOQ switches every input provides a single and separate FIFO for each output. Such a FIFO is called a Virtual Output Queue. When a new cell arrives at the input port, it is stored in the destined queue and waits for transmission through a switching fabric.

To solve internal blocking and output port contention issues in VOQ switches fast arbitration schemes are needed. An arbitration scheme is essentially a service discipline that arranges the transmission order among the input cells. It decides which items of information should be passed from inputs to arbiters, and – based on that decision – how each arbiter picks one cell from among all input cells destined for the output. The arbitration decisions for every output port have to be taken in each time slot using a central arbiter, or distributed arbiters. In the distributed manner, each output has its own arbiter operating independently from others. However, in this case it is necessary to send many request-grant-accept signals.

It is very difficult to implement such arbitration in the real environment because of time constraints. A central arbiter may also create a bottleneck due to time constraints as the switch size increases.
Considerable work has been done on scheduling algorithms for the crossbar and three-stage Clos-network VOQ switches. Most of them achieve 100% throughput under the uniform traffic, but the throughput is usually reduced under the nonuniform traffic (Chao & Liu, 2007). A switch can achieve 100% throughput under the uniform or nonuniform traffic if the switch is stable, as it was defined in (McKeown at al., 1999). In general, a switch is stable for a particular arrival process if the expected length of the input queues does not grow without limits.
This chapter presents basic ideas concerning packet switching in next generation switches/routers. The simulation results obtained by us for the well known and new packet dispatching schemes for the three-stage buffered Clos-network switches are also shown and discussed. The remainder of the chapter is organized as follows: subchapter 2 introduces some background knowledge concerning the Clos-network switch that we refer to throughout this chapter; subchapter 3 presents packet dispatching schemes with distributed arbitration; subchapter 4 is devoted to dispatching schemes with centralized arbitration. A survey of related works is carried out in subchapter 5.

2. Clos switching network

In 1953, Clos proposed a class of space-division three-stage switching networks and proved strictly non-blocking conditions of such networks (Clos, 1953). These kind of switching fabrics are widely used and extensively studied as a scalable and modular architecture for the next generation switches/routers. The Clos switching fabric can achieve a nonblocking property with the smaller number of total crosspoints in the switching elements than crossbar switches. Nonblocking switching fabrics are divided into four classes: strictly nonblocking (SSNB), wide-sense nonblocking (WSNB), rearrageable nonblocking (RRNB) and repackably nonblocking (RPNB) (Kabacinski, 2005). SSNB and WSNB ensures, that any pair of idle input and output can be connected without changing any existing connections, but a special path set-up strategy must be used in WSNB networks. In RRNB and RPNB any such pair can be also connected, but it may be necessary to re-switch existing connections to other connecting paths. The difference is in time these reswitchings take place. In RRNB, when a new request arrives, and is blocked, an appropriate control algorithm is used to reswitch some of existing connections to unblock the new call. In RPNB, a new call can always be set up without reswitching of existing connections, but reswitching takes place when any of existing call is terminated. These reswitchings are done to prevent a switching fabric from blocking states before a new connection arrives.
The three-stage Clos-network architecture is denoted by $C(m, n, k)$, where parameters m, n, and k entirely determine the structure of the network. There are k input switches of capacity $n \times m$ in the first stage, m switches of capacity $k \times k$ in the second stage, and k output switches of capacity $m \times n$ in the third stage. The capacity of this switching system is $N \times N$, where $N = nk$. The three-stage Clos switching fabric is strictly nonblocking if $m \geq 2n-1$ and rearrangeable nonblocking if $m \geq n$. The three-stage Clos-network switch architecture may be categorized into two types: bufferless and buffered. The former one has no memory in any stage, and it is also referred to as the Space-Space-Space (S^3) Clos-network switch, while

the latter one employs shared memory modules in the first and third stages, and is referred to as the Memory-Space-Memory (MSM) Clos-network switch. The buffers in the second stage modules cause an out-of-sequence problem, so a re-sequencing function unit in the third stage modules is necessary but difficult to implement when the port speed increases. One disadvantage of the MSM architecture is that the first and third stages are both composed of shared-memory modules.

We define the MSM Clos switching fabric based on the terminology used in (Oki at al., 2002a) (see Fig. 4 and Table 1).

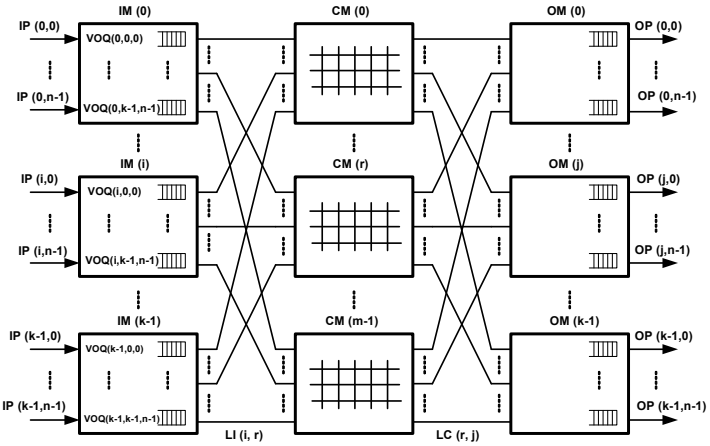


Fig. 4. The MSM Clos switching network

Notation	Description
IM	Input module at the first stage
CM	Central module at the second stage
OM	Output module at the third stage
i	IM number, where $0 \le i \le k-1$
j	OM number, where $0 \le j \le k-1$
h	Input/output port number in each IM/OM, where $0 \le h \le n-1$
r	CM number, where $0 \le r \le m-1$
IM (i)	The *(i+1)*th input module
CM (r)	The *(r+1)*th central module
OM (j)	The *(j+1)*th output module
IP (i, h)	The *(h+1)*th input port at *IM(i)*
OP (j, h)	The *(h+1)*th output port at *OM(j)*
LI (i, r)	Output link at *IM(i)* that is connected to *CM(r)*
LC (r, j)	Output link at *CM(r)* that is connected to *OM(j)*
VOQ (i, j, h)	Virtual output queue that stores cells from *IM(i)* to *OP(j, h)*

Table 1. A notation for the MSM Clos switching fabric

In the MSM Clos switching fabric architecture the first stage consists of k IMs, and each of them has an $n \times m$ dimension and nk VOQs to eliminate Head-Of-Line blocking. The second stage consists of m bufferless CMs, and each of them has a $k \times k$ dimension. The third stage

consists of k OMs of capacity $m \times n$, where each *OP(j, h)* has an output buffer. Each output buffer can receive at most m cells from m CMs, so a memory speedup is required here.
Generally speaking, in the MSM Clos switching fabric architecture each *VOQ(i, j, h)* located in *IM(i)* stores cells going from *IM(i)* to *OP(j, h)* at *OM(j)*. In one cell time slot VOQ can receive at most n cells from n input ports and send one cell to any CMs. A memory speedup of n is required here, because the rate of memory work has to be n times higher than the line rate. Each *IM(i)* has m output links connected to each *CM(r)*, respectively. A *CM(r)* has k output links *LC(r, j),* which are connected to each *OM(j)*, respectively.
Input buffers located in IMs may be also arranged as follows:

- An input buffer in each input port is divided into N parallel queues, each of them storing cells directed to different output ports. Each IM has nN VOQs, no memory speedup is required.
- An input buffer in each IM is divided into k parallel queues, each of them storing cells destined to different OMs. Those queues will be called Virtual Output Module Queues (VOMQs), instead of VOQs. It is possible to arrange buffers in such way because OMs are nonblocking. Memory speedup of n is necessary here. In that case, there are less queues in each IMs but they are longer than VOQs. Each *VOMQ(i, j)* stores cells going from *IM(i)* to the *OM(j).*
- Each input of an IM has k parallel queues, each of them storing cells destined to different OMs; we call it mVOMQs (multiple VOMQs). In each IM there are nk mVOMQs. This type of buffer arrangement eliminates a memory speedup. Each *mVOMQ(i, j, h)* stores cells going from *IP(i, h)* to the *OM(j)*, h denotes the input port number or the number of a VOMQ group.

Thanks to allocating buffers in the first and third stages the main switching problem in the three-stage buffered Clos-network switches lies in routes assignment between input and output modules.

3. Packet dispatching algorithms with distributed arbitration

The packet dispatching algorithms are responsible for choosing cells to be sent from the VOQs to the output buffers, and simultaneously for selecting connecting paths from IMs to OMs. Considerable work has been done on packet dispatching algorithms for the three-stage buffered Clos-network switches. Unfortunately, the known optimal algorithms are too complex to implement at very high data rates, so sub-optimal, heuristic algorithms of lesser complexity, but also lesser performance, have to be used. The idea of three-phase algorithm, namely request-grant-accept, described by Hui and Arthurs (Hui & Arthus, 1987), is widely used by the packet dispatching algorithms with distributed arbitration. In this algorithm many request, grant and accept signals are sent between each input and output to do matching. In general, the three-phase algorithm works as follows: each unmatched input sends a request to every output for which it has a queued cell. If an unmatched output receives multiple requests, it grants one over all requests. If an input receives multiple grants, it accepts one and sends an accept signal to matched output. These three steps may be repeated in many iterations.
The primary multiple-phase dispatching algorithms for the three-stage buffered Clos-network switches were proposed in (Oki at al. 2002a). The basic idea of these algorithms is to use the effect of desynchronization of arbitration pointers and common request-grant-

accept handshaking scheme. The well known algorithm with multiple-phase iterations is the CRRD (Concurrent Round-Robin Dispatching). Other algorithms like the CMSD (Concurrent Master-Slave Round-Robin Dispatching) (Oki at al. 2002a), SRRD (Static Round-Robin Dispatching) (Pun & Hamdi, 2004), and proposed by us in (Kleban & Wieczorek, 2006) - CRRD-OG (Concurrent Round-Robin Dispatching with Open Grants) use the main idea of the CRRD scheme and try to improve results by implementing different mechanisms. We start to describe these algorithms with presentation of very simple scheme called Random Dispatching (RD).

3.1 Random dispatching scheme

Random selection as dispatching scheme is used by the ATLANTA switch developed by Lucent Technologies (Chao & Liu, 2007). An explanation of the basic concept of Random Dispatching (RD) scheme should help us to understand how the CRRD and CRRD-OG algorithms work.

The basic idea of RD scheme is quite similar to the PIM (Parallel Iterative Matching) scheduling algorithm used in the single stage switches. In this scheme two phases are considered for dispatching from the first to second stages. In the first phase each IM randomly selects up to m VOQs and assigns them to IM output links. In the second phase requests associated with output links are sent from an IM to a CM. The arbitration results are sent from CMs to IMs, so the matching between IMs and CMs can be completed. If there is more than one request for the same output link in the CM, it grants one request randomly. In the next time slot the granted VOQs will transfer their cells to the corresponding OPs.

In detail, the RD algorithm works as follows:

- *PHASE 1*: Matching within IM:
  - *Step 1:* Each nonempty VOQ sends a request for candidate selection.
  - *Step 2*: The *IM(i)* selects up to m requests out of nk nonempty VOQs. A round-robin arbitration can be employed for this selection.
- *PHASE 2:* Matching between IM and CM:
  - *Step 1:* A request that is associated with *LI(i, r)* is sent out to the corresponding *CM(r)*. An arbiter that is associated with *LC(r, j)* selects one request among k and the *CM(r)* sends up to k grants, each of which is associated with one *LC(r, j)*, to the corresponding IMs.
  - *Step 2:* If the VOQ at the IM receives the grant from the CM, it sends the corresponding cell at the next time slot. Otherwise, the VOQ will be a candidate again at step 2 in Phase 1 at the next time slot.

It has been shown that a high switch throughput cannot be achieved due to the contention at the CM, unless the internal bandwidth is expanded. To achieve 100% throughput the expansion ratio m/n has to be set to at least: $(1–1/e)^{-1} \approx 1{,}582$ (Oki at al. 2002a).

3.2 Concurrent Round-Robin Dispatching

The Concurrent Round Robin Dispatching (CRRD) algorithm has been proposed to overcome the throughput limitation of the RD scheme. The basic idea of this algorithm is to use the desynchronization of arbitration pointers effect in the three-stage Clos-network switch. It is based on common request-grant-accept handshaking scheme and achieves 100%

throughput under uniform traffic. To easily obtain pointers desynchronization effect the *VOQ(i, j, h)* in the *IM(i)* are rearranged for dispatching as follows:

VOQ(i, 0, 0), VOQ(i, 1, 0), VOQ(i, 2, 0), ... , VOQ(i, k-1, 0)
VOQ(i, 0, 1), VOQ(i, 1, 1), VOQ(i, 2, 1), ... , VOQ(i, k-1, 1)
...
VOQ(i, 0, n-1), VOQ(i, 1, n-1), VOQ(i, 2, n-1) ,..., VOQ(i, k-1, n-1)
Therefore, *VOQ(i, j, h)* is redefined as *VOQ(i, v)*, where $v = hk + j$ and $0 \leq v \leq nk - 1$.
Each *IM(i)* has *m* output link round-robin arbiters and *nk* VOQ round-robin arbiters. An output link arbiter associated with *LI(i, r)* has its own pointer *PL(i, r)*. A VOQ arbiter associated with the *VOQ(i, v)* has its own pointer *PV(i, v)*. In *CM(r)*, there are *k* round robin arbiters, each of which corresponds to *LC(r, j)* – an output link to the *OM(j)* – and has its own pointer *PC(r, j)*.
The CRRD algorithm completes the matching process in two phases. In Phase 1 at most *m* VOQs are selected as candidates, and the selected VOQ is assigned to an IM output link. An iterative matching with round-robin arbiters is adopted within the *IM(i)* to determine the matching between a request from the *VOQ(i, v)* and the output link *LI(i, r)*. This matching is similar to the iSLIP approach (Chao & Liu, 2007). In Phase 2, each selected VOQ that is associated with each IM output link sends a request from an IM to a CM. The CMs respond with the arbitration results to IMs so that the matching between IMs and CMs can be done.
The pointers *PL(i, r)* and *PV(i, v)* in the *IM(i)* and *PC(r, j)* in the *CM(r)* are updated to one position after the granted position, only if the matching within the IM is achieved at the first iteration on Phase 1 and the request is also granted by the CM in Phase 2.
It was shown that there is a noticeable improvement in the cell average delay by increasing the number of iterations in each IM. However, the number of iterations is limited by the arbitration time in advance. Simulation results obtained by us shown that the optimal number of iterations in the IM is *n*/2 and more iterations do not produce a measurable improvement.
The CRRD algorithm works as follows:

- *PHASE 1:* Matching within IM
  First iteration:
  - *Step 1:* Request: Each nonempty *VOQ(i, v)* sends a request to every arbiter of the output link *LI(i, r)* within *IM(i)*.
  - *Step 2:* Grant: Each arbiter of the output link *LI(i, r)* chooses one VOQ request in a round-robin fashion and sends the grant to the selected VOQ. It starts searching from the position of *PL(i, r)*.
  - *Step 3:* Accept: Each arbiter of *VOQ(i, v)* chooses one grant in a round-robin fashion and sends the accept to the matched output link *LI(i, r)*. It starts searching from the position of *PV(i, v)*.

  i-th iteration (*i*>1):
  - *Step 1:* Each unmatched *VOQ(i, v)* at the previous iterations sends another request to all unmatched output link arbiters.
  - *Step 2 and 3:* These steps are the same as in the first iteration.
- *PHASE 2:* Matching between IM and CM
  - *Step 1:* Request: Each selected in Phase 1 IM output link *LI(i, r)* sends the request to *CM(r) j*th output link *LC(r, j)*.

- *Step 2:* Grant: Each round-robin arbiter associated with output link *LC(r, j)* chooses one request by searching from the position of *PC(r, j),* and sends the grant to the matched output link *LI(i, r)* of *IM(i).*
- *Step 3:* Accept: If the *LI(i, r)* receives the grant from the *LC(r, j)* it sends the cell from the matched *VOQ(i, v)* to the *OP(j, h)* through the *CM(r)* at the next time slot. The IM cannot send the cell without receiving the grant. Not granted requests from the CM will be again attempted to be matched at the next time slot because the round-robin pointers are updated to one position after the granted position only if the matching within IM is achieved in Phase 1 and the request is also granted by the CM in Phase 2.

3.3 Concurrent Round-Robin Dispatching with Open Grants

The Concurrent Round-Robin Dispatching with Open Grants (CRRD-OG) algorithm is an improved version of the CRRD scheme in terms of the number of iterations which are necessary to achieve better results. In the CRRD-OG algorithm a mechanism of open grants is implemented. An open grant is sent by a CM to an IM and contains information about unmatched link from the second to the third stage. In other words, the *IM(i)* is informed about unmatched output link *LC(r, j)* to the *OM(j).* The open grant is sent by each unmatched output link *LC(r, j)*. Due to the architecture of the three-stage Clos switching fabric is clearly defined, it is also information about output port numbers, which can be reached using the output *j* of the *CM(r).* On the basis of this information the *IM(i)* looks up through VOQs and searches a cell which is destined to any output of the *OM(j)*. If such cell exists it will be sent at the next time slot. To support the process of searching the proper cell to be sent to the *OM(j)* each IM has *k* open grant arbiters with *POG(i, j)* pointers. Each arbiter is associated with the *OM(j)* accessible by the output link *LC(r, j)* of the *CM(r*). The *POG(i, j)* pointer is used to search VOQs located at each input port according to the round robin routine.

In the CRRD-OG algorithm two phases are necessary to complete matching process. Phase 1 is the same as in the CRRD algorithm. In Phase 2 the CRRD-OG algorithm works as follows:

- *PHASE 2:* Matching between IM and CM
  - *Step 1:* Request: Each selected in Phase 1 IM output link *LI(i, r)* sends the request to the *CM(r) j*th output link *LC(r, j).*
  - *Step 2:* Grant: Each round-robin arbiter associated with the output link *LC(r, j)* chooses one request by searching from the position of *PC(r, j),* and sends the grant to the matched *LI(i, r)* of *IM(i).*
  - *Step 3:* Open Grant: If after step 2, the unmatched output links *LC(r, j)* still exist, each unmatched output link *LC(r, j)* sends the open grant to the output link *LI(i, r)* of the *IM(i).* The open grant contains the idle output's number of the CM module, which simultaneously determine the *OM(j)* and accessible outputs of the Clos switching fabric.
  - *Step 4:* If the *LI(i, r)* receives the grant from the *LC(r, j)* it sends the cell, at the next time slot, from the matched *VOQ(i, v)* to the *OP(j, h)* through the *CM(r).* If the *LI(i, r)* receives the open grant from the *LC(r, j)* the open grant arbiter has to choose one cell, which is destined to *OM(j)* and sends it at the next time slot. The open grant arbiter starts to go through the VOQs looking for the proper cell from the position shown by

the *POG(i, k)* pointer. The IM cannot send the cell without receiving the grant or the open grant. Not granted requests will be again attempted to be matched at the next time slot because the pointers are updated only if the matching is achieved. If the cell is sent as a reaction to the open grant the pointers are updated under the following conditions:

- if the pointer *PL(i, r)* points the VOQ which sent the cell, it is updated;
- if the pointer *PV(i, v)* points the output used to sent the cell, it is updated;
- if the pointer *PC(r, j)* points the link *LI(i, r)* used to sent the open grant, it is updated.

Fig. 5-10 illustrates the details of the CRRD-OG algorithm by showing an example for the Clos network *C(3, 3, 3)*.

- *PHASE 1:* Matching within *IM(2)* (one iteration).
  - *Step 1:* The nonempty VOQs: *VOQ(2, 0), VOQ(2, 2), VOQ(2, 3), VOQ(2, 4)*, and *VOQ(2, 8)* send requests to all output link arbiters (Fig. 5).

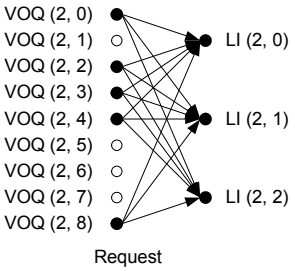


Fig. 5. Nonempty VOQs send requests to all output link arbiters

  - *Step 2:* Output link arbiters associated with *LI(2, 0), LI(2, 1)* and *LI(2, 2)* select *VOQ(2, 0), VOQ(2, 2)* and *VOQ(2, 3)* respectively, according to their pointers position and send grants to them (Fig. 6).

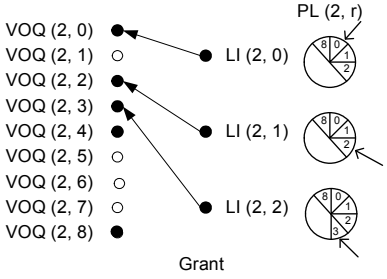


Fig. 6. Output link arbiters send grants to selected VOQs

  - Step 3. Each selected VOQ: *VOQ(2, 0), VOQ(2, 2)* and *VOQ(2, 3)*, receives only one grant, and sends accept to the proper output link arbiter (Fig. 7).

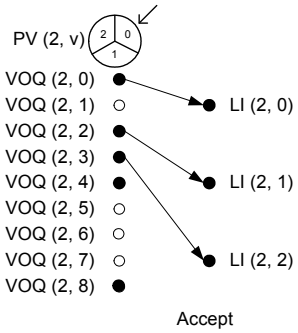


Fig. 7. VOQs send accept to chosen output link arbiters

- PHASE 2: Matching between IM and CM (as an example we consider the state in *CM(2)*).
  - Step 1. In this step the output links of *CM(2)* receive requests from the output links of IMs matched in Phase 1. The requests are as follows: *LC(2, 0), LC(2, 1), LC(2, 0)* (Fig. 8).

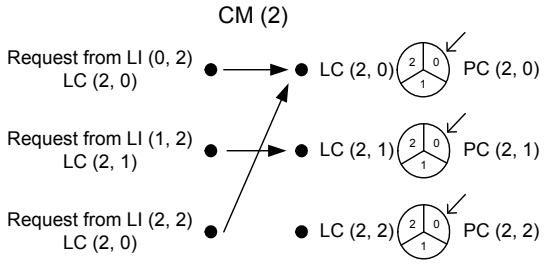


Fig. 8. Output link arbiters of the CM(2) receive requests

  - Step 2. The output link arbiter *LC(2, 0)* receives two requests from *IM(0)* and *IM(2)*, and selects the request from *IM(0)*, according to the pointer position. The output link arbiter *LC(2, 1)* selects request from *IM(2)*. Output links arbiters: *LC(2, 0)* and *LC(2, 1)* send grants to *IM(0)* and *IM(1)* respectively.
  - Step 3. The output link arbiter *LC(2, 2)* does not receive a request, so it sends open grant to *IM(2)* (Fig. 9).

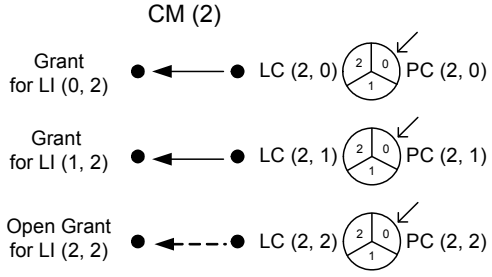


Fig. 9. The output port arbiter LC(2, 2) sends the open grant to LI (2, 2)

- Step 4. *IM(2)* receives the open grant from *LC(2, 2),* which means that it is possible to send one cell to *OP(2, h)*. It chooses a cell from *VOQ(2, 8)*. The cell is destined to *OP(2, 2)* (Fig. 10), and is sent at the next time slot, together with other cells from IMs to OMs through CMs.

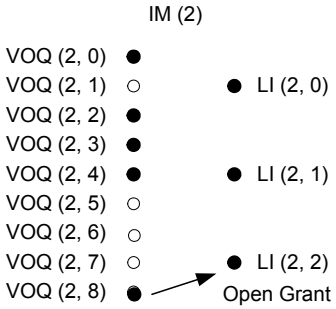


Fig. 10. The cell from VOQ(2, 8) is matched with LI(2, 2), as a reaction to the open grant received from LC(2, 2)

3.4 Concurrent Master-Slave Round-Robin Dispatching Scheme

The Concurrent Master-Slave Round-Robin Dispatching (CMSD) algorithm is an improved version of the CRRD algorithm. It preserves all advantages of the CRRD scheme but more arbiters are used to perform the iterative matching process within the IMs. Two sets of round-robin arbiters (master and slave) are employed to perform hierarchical round-robin arbitration process in the first stage of the Clos-network switch. Each output link of IMs is associated with one master and *k* slaves arbiters. To describe the CMSD algorithm we define several notations based on the terminology used in (Oki at al., 2002a). A VOQ group that consists of *n* VOQs storing cells from *IM(i)* to *OM(j)* is denoted by *G(i, j)*. Each IM has *m* master output-link round robin arbiters, denoted as *ML(i, r), mk* slave output-link round-robin arbiters, denoted as *SL(i, j, r),* and *nk* VOQ round-robin arbiters. Each master arbiter associated with *LI(i, r)* has its own pointer *PML(i, r)*. Each slave arbiter associated with *LI(i, r)* and *G(i, j)* has its own pointer *PSL(i, j, r)*. Each VOQ arbiter associated with *VOQ(i, j, h)* has its own pointer *PV(i, j, h).* The master arbiter is responsible for selection of one nonempty *G(i, j)* group, while the slave arbiter selects one nonempty VOQ within that VOQ group.

The CMSD algorithm works as follows:

- *PHASE 1:* Matching within IM

  First iteration:

  - *Step 1:* Request: Two sets of requests are sent to the output link arbiters. The group-level request are sent from *G(i, j)* that has at least one non-empty VOQ to every master arbiter *ML(i, r)*. At the same time, each nonempty *VOQ(i, j, h)* sends a request to every slave arbiter *SL(i, j, r)*.
  - *Step 2:* Grant: Each round-robin master arbiter *ML(i, r)* chooses a request by searching from the position of *PML(i, r)*. At the same time, each slave arbiter selects one VOQ request in a round-robin fashion by searching from the position of *PSL(i, j, r)*. The slave arbiter *SL(i, j, r)* will send the grant to the selected VOQ only if *G(i, j)* has been

selected by its master arbiter. If *SL(i, j, r)* does not receive a grant, the search is invalid.

- *Step 3:* Accept: Each VOQ arbiter searches in a round-robin fashion one grant, among all those received, and sends the accept to the master and slave output-link arbiters. Each arbiter starts searching a grant from the position of *PV(i, j, h)*.

i-th iteration (*i>1*):

- *Step 1:* Each *VOQ(i, j, h)* unmatched at the previous iterations sends another request to the slave arbiters. The group *G(i, j)*, which has at least one unmatched nonempty VOQ sends a request to the master arbiters.
- *Step 2 and 3:* These steps are the same as in the first iteration.

➢ *PHASE 2:* Matching between IM and CM – the matching procedure is the same as in the CRRD algorithm.

All the round-robin pointers located in *IM(i)* (namely: *PML(i, r), PSL(i, j, r)* and *PV(i, j, h))* and in *CM(r)* (namely: *PC(r, j))* are updated to one position after the granted position only if the matching is achieved at Phase 1 and the request is also granted at Phase 2. The CMSD algorithm can very easily achieve the desynchronization effect of all round-robin pointers, so it can provide high throughput without expansion under the uniform traffic.

3.5 Static Round-Robin Dispatching

The Static Round-Robin Dispatching (SRRD) algorithm was proposed by K. Pun and M. Hamdi, and is an adaptation of the Static Round-Robin (SRR) scheme for the MSM Clos-network switches. The SRR algorithm was first introduced by Jiang and Hamdi in (Jiang & Hamdi, 2001) for crossbar switches and uses the same handshaking scheme as in the iSlip or DRRM scheme (Chao & Liu, 2007). The algorithm is simple and can achieve very good delay performance. In this algorithm the arbitration pointers are artificially set to be desynchronized at the beginning, and are updated in a static way to keep them desynchronized all the time. Additionally, the grant and accept pointers are "mutual matched". That is, if grant pointer g_j in output j is pointing to input i, then accept pointer a_i in input i must point to output j. The matching sequence in SRR scheme is shown in Fig. 11. This allows the maximum matching from input ports to output ports if all VOQs have a cell to be sent.

The SRRD scheme works in the same way as the CMSD scheme, which means that the phases and steps are in both algorithms identical except the pointer initialization and pointer updating. The initial values of the pointers are as follows: *PV(i, j, h)=h, PSL(i, j, r)=r, PML(i, r)=(i+r)%k* and *PC(r, j)=i if PML(i, r)=j*. The pointers *PML(i, r)* and *PC(r, j)* are always incremented by one (mod k), but the pointers *PV(i, j, h)* and *PSL(i, j, r)* remain unchanged, no matter there is a match or not.

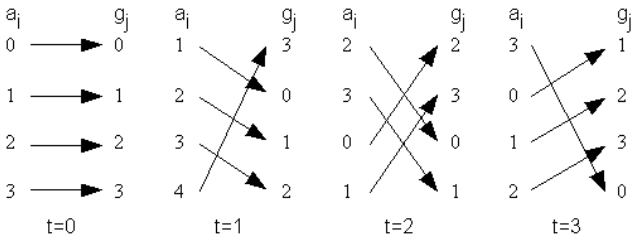


Fig. 11. Matching sequence in SRR algorithm

The SRRD scheme can always achieve 100% throughput under the uniform traffic. Unfortunately, due to several arbiters may grant the same request at the same time, the performance under nonuniform traffic is degraded. This phenomenon appears because all conventional arbiters search in clock-wise direction. To improve the performance of the MSM Clos switch under the nonuniform traffic distribution patterns it is necessary to allow some round-robin arbiters to search the requests in clockwise direction and anti-clockwise direction alternatively, each for one time slot. The 0/1 counter is necessary to keep track of time. The counter is incremented by one (mod 2) in each time slot. If counter shows 0 the master arbiter *ML(i, r)* searches one request in clockwise round-robin fashion, otherwise if counter shows 1, the master arbiter searches one request in anti-clockwise round-robin fashion.

3.6 Performance of CRRD, CMSD, SRRD and CRRD-OG algorithms

A. Packet Arrival Models

Two packet arrival models namely the Bernoulli and bursty are considered in simulation experiments. In the Bernoulli arrival model cells arrive at each input in slot-by-slot manner and the probability that there is a cell arriving in each time slot is identical and independent of any other slot. The probability that a cell may arrive in a time slot is denoted by p and is referred to as the load of the input. This type of traffic defines a memoryless random arrival pattern.

In the bursty traffic model, each input alternates between active and idle periods. During active periods, cells destined for the same output arrive continuously in consecutive time slots. The average burst (active period) length is set to 16 cells in our simulations.

B. Traffic distribution models

We consider several traffic distribution models which determine the probability that a cell which arrives at an input will be directed to a certain output. The considered traffic models are:

Uniform traffic – this type of traffic is the most commonly used traffic profile. In the uniformly distributed traffic probability p_{ij} that a packet from input i will be directed to output j is uniformly distributed through all outputs, i.e.:

$$p_{ij} = p/N \; \forall i, j \tag{1}$$

Trans-diagonal traffic – in this traffic model some outputs have a higher probability of being selected, and respective probability p_{ij} was calculated according to the following equation:

$$p_{ij} = \begin{cases} \frac{p}{2} & for \;\; i = j \\ \frac{p}{2(N-1)} & for \;\; i \neq j \end{cases} \tag{2}$$

Bi-diagonal traffic – is very similar to the trans-diagonal traffic but packets are directed to one of two outputs, and respective probability p_{ij} was calculated according to the following equation:

$$p_{ij} = \begin{cases} \frac{2}{3}p & for\ i = j \\ \frac{p}{3} & for\ j = (i+1) \mod N \\ 0 & otherwise \end{cases} \quad (3)$$

Chang's traffic – this model is defined as:

$$p_{ij} = \begin{cases} 0 & for\ i = j \\ \frac{1}{N-1} & otherwise \end{cases} \quad (4)$$

The experiments have been carried out for the MSM Clos switching fabric of size 64 × 64 - C(8, 8, 8), and for a wide range of traffic load per input port: from p = 0.05 to p = 1, with the step 0.05. The 95% confidence intervals that have been calculated after t-student distribution for ten series, per 55000 cycles each (after the starting phase comprising 15000 cycles, which enables to reach the stable state of the switching fabric), are at least one order lower than the mean value of the simulation results, therefore they are not shown in the figures. We have evaluated two performance measures: the average cell delay in time slots and the maximum VOQs size for the CRRD, CMSD, SRRD, and CRRD-OG algorithms. The results of the simulation under 1 and/or 4 iterations (represented in figures by itr) are shown in the charts (Fig. 12-21). In any case, the number of iterations between any IM and CM is one.

Fig. 12, 14, 16, 18 show the average cell delay in time slots obtained for the uniform, Chang's, trans-diagonal and bi-diagonal traffic patterns, whereas Fig. 13, 15, 17, 19 show the maximum VOQ size in a number of cells. To make the charts more clear and lucid only results for itr=4 are shown in figures concerning the maximum VOQ size. Fig. 20 and 21 show the results for the bursty traffic with the average burst length set to 16 cells.

We can observe that using the Bernoulli traffic and all investigated traffic distribution patterns the CRRD-OG algorithm provides better performance than the CRRD, CMSD and SRRD algorithms. In many cases the CRRD-OG algorithm with one iteration delivers better performance than other algorithms with four iterations (see Fig. 12, 14, 16). The same relation between the CRRD-OG scheme and others schemes we can notice under the bursty traffic (Fig. 20).

Under the uniform traffic the SRRD scheme gives only slightly worse results than the CRRD-OG scheme; the worst result gives pure CRRD algorithm. The same relation we can see in Fig. 13 which shows the comparison of the maximum VOQ size. The biggest buffers we need if we control the MSM Clos-network switch using the CRRD algorithm. The Chang's distribution traffic pattern is very similar to the uniform distribution traffic pattern. Under this traffic distribution pattern all algorithms receive 100% throughput and CRRD-OG scheme with one iteration delivers better performance than other algorithms with four iterations for the cell delay as well as the maximal VOQ size. (Fig. 14, 15). The trans-diagonal and bi-diagonal traffic distribution patterns are highly demanding and the investigated packet dispatching schemes cannot provide the 100% throughput for the MSM Clos – network switch. The best results have been obtained for the CRRD-OG scheme with 4 iterations. These are respectively: under trans-diagonal traffic pattern - 80% throughput for one iteration and 85% throughput for four iterations (Fig. 16) and under bi-diagonal traffic pattern – 95% (Fig. 18). Under the bursty packet arrival model the CRRD-OG scheme

provides much better performance than other algorithms especially for the very high input load (Fig. 20). The same relationship as for the cell delay we can observe for the maximal VOQs size (Fig. 13, 15, 17, 19, 21). It is obvious that for small cell delay the size of VOQs will be also small.

The simulation experiments have shown that the CRRD-OG scheme with one iteration gives very good results in the average cell delay and VOQs size. An increase in the number of iterations do not produce further significant improvement, quite the opposite to other iterative algorithms. Particularly more than n/2 iterations do not change significantly the performance of all investigated iterative schemes.

The investigated packet dispatching schemes are based on the effect of desynchronization of arbitration pointers in the Clos-network switch. In our research we have made an attempt to improve the method of pointers desynchronization for the CRRD-OG scheme, to ensure the 100% throughput for the nonuniform traffic distribution patterns. Additional pointers and arbiters for open grants had been added to the MSM Clos-network switch, but the scheme was not able to provide 100% throughput for the nonuniform traffic distribution patterns. To our best knowledge it is not possible to achieve very good desynchronization of pointers using the methods implemented in the iterative packet dispatching schemes. In our opinion the decisions of the distributed arbiters have to be supported by the central arbiter, but the implementation of such solution in the real equipment will be very complex.

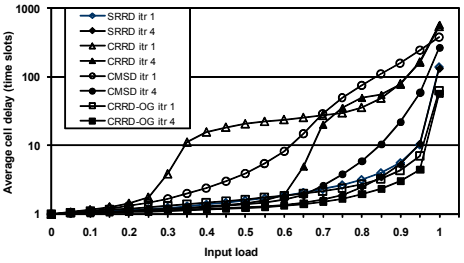


Fig. 12. Average cell delay, uniform traffic

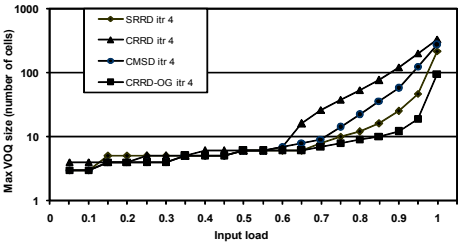


Fig. 13. Maximum VOQ size, uniform traffic

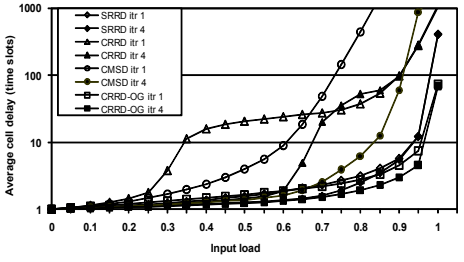


Fig. 14. Average cell delay, Chang's traffic

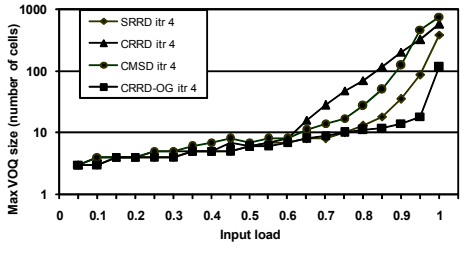


Fig. 15. Maximum VOQ size, Chang's traffic

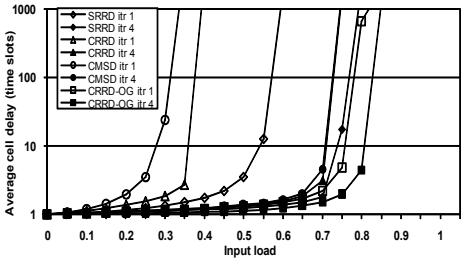


Fig. 16. Average cell delay, trans-diagonal traffic

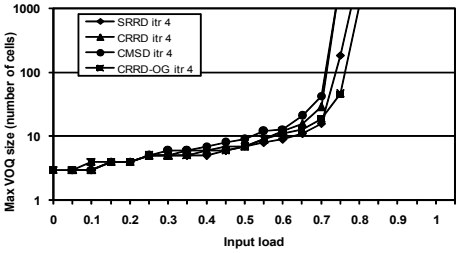


Fig. 17. Maximum VOQ size, trans-diagonal

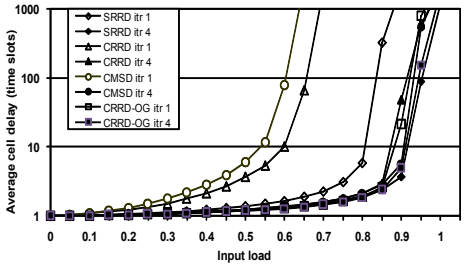


Fig. 18. Average cell delay, bi-diagonal traffic

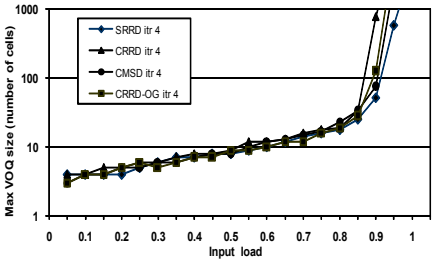


Fig. 19. Maximum VOQ size, bi-diagonal traffic

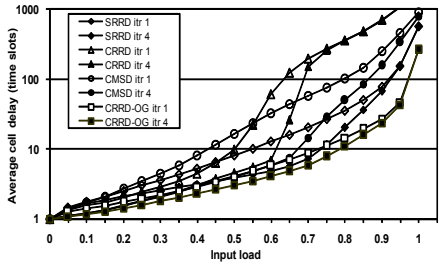


Fig. 20. Average cell delay, bursty traffic, average burst length b=16

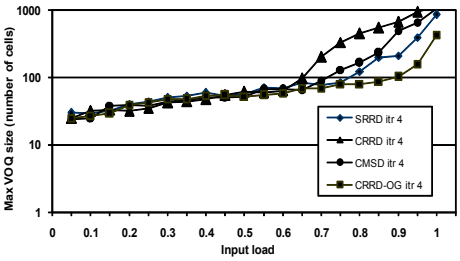


Fig. 21. Maximum VOQ size, bursty traffic, average burst length b=16

4. Packet dispatching algorithms with centralized arbitration

The packet dispatching algorithms with centralized arbitration use a central arbiter to take packet scheduling decisions. Currently, the central arbiters are used to control one-stage switching fabrics. This subchapter presents three packet dispatching schemes with centralized arbitration for the MSM Clos-network switches. We call these schemes as follows: Static Dispatching-First Choice (SD-FC), Static Dispatching-Optimal Choice (SD-OC) and Input Module - Output Module Matching (IOM).

Packet switching nodes in the next generation Internet should be ready to support the nonuniform/hot spot traffic. Such case often occurs when a popular server is connected to a single switch/router port. Under the nonuniform traffic distribution patterns selected VOQs store more cells than others. Due to some input buffers may be overloaded, it is necessary to implement to a packet dispatching scheme a special mechanism, which is able to send up to n cells from *IM(i)* to *OM(j)* in the same time slot, in order to unload overloaded buffers. Three dispatching schemes presented in this subchapter have such possibility.

The SD-FC, SD-OC, and IOM schemes make a matching between each IM and OM, taking into account the number of cells waiting in VOMQs. Each VOMQ has its own counter *PV(i, j),* which shows the number of cells destined to *OM(j)*. The value of *PV(i, j)* is increased by 1 when a new cell is written into a memory, and decreased by 1 when a cell is sent out to *OM(j)*. The algorithms use the central arbiter to indicate the matched pairs of *IM(i)-OM(j)*. The set of data sent to the arbiter by each scheme is different, therefore, the architecture and functionality of each arbiter is also different. After a matching phase, in the next time slot *IM(i)* is allowed to send up to n cells to the selected *OM(j).*

In the SD-OC and SD-FC schemes the central arbiter matches *IM(i)* and *OM(j)* only if the number of cells buffered in *VOMQ(i, j)* is at least equal to n. Under the nonuniform traffic distribution patterns it happens very often, contrary to the uniform traffic distribution. In the proposed packet dispatching schemes each VOMQ has to wait until at least n cells are stored before being allowed to make a request. In simulation experiments we consider the Clos switching fabric without any expansion, denoted by *C(n, n, n)*, so in description of the packet dispatching schemes, k and m parameters are not used.

4.1 Static Dispatching

To reduce latency and avoid starvation, a very simple packet dispatching routine, called Static Dispatching (SD), is also used in the MSM Clos-network switch to support SD-FC and SD-OC schemes. Under this algorithm, connecting paths in switching fabric are set up according to static, but different in each CM, connection patterns (see Fig. 22). These fixed connection paths between IMs and OMs eliminate the handshaking process with the second stage, and no internal conflicts in the switching fabric will occur. Also no arbitration process is necessary. Cells destined to the same OM, but located in different IMs, will be sent through different CMs.

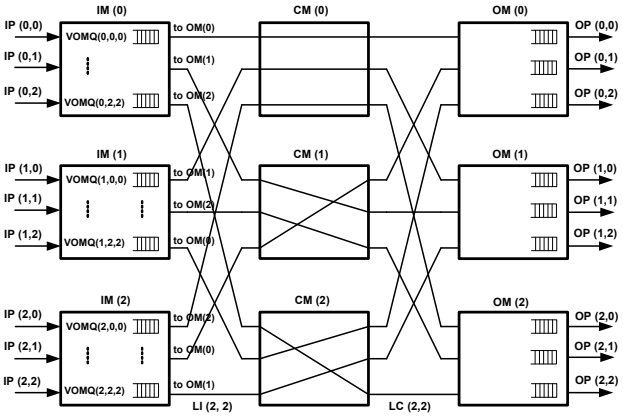


Fig. 22. Static connection patterns in CMs, *C(3, 3, 3)*.

In detail, the SD algorithm works as follows:

- *Step 1:* According to the connection pattern of *IM(i)*, match all output links *LI(i, r)* with cells from VOMQs.
- *Step 2:* Send the matched cells in the next time slot. If there is any unmatched output link, it remains idle.

4.2 Static Dispatching-First Choice and Static Dispatching-Optimal Choice Schemes

The SD-OC and SD-FC schemes are very similar, but the central arbiter matching IMs and OMs works in a different way. In both algorithms the *PV(i, j)* counter, which reaches the value equal or greater than *n* sends the information about an overloaded buffer to the central arbiter. In the central arbiter there is a binary matrix representing VOMQs load. If the value of matrix element *x[i, j]*=1, it means that *IM(i)* has at least *n* cells that should be sent to *OM(j)*.

In the SD-OC scheme the main task of the central arbiter is to find an optimal set of 1s in the matrix. The best case is *n* 1s, but it is possible to choose only single 1 from column *i* and row *j*. If there is no such set of 1s the arbiter tries to find a set of *n-1* 1s, which fulfills the same conditions, and so on. The round-robin routine is used for the starting point of the searching process. Otherwise, the MSM Clos switching fabric is working under the SD scheme.

The main difference between the SD-OC and SD-FC lies in the operation of the central arbiter. In the SD-FC scheme the central arbiter does not look for the optimal set of 1s, but

tries to match *IM(i)* with *OM(j)*, choosing the first 1 found in column *i* and row *j*. No optimization process for selecting IM-OM pairs is employed. In detail, the SD-OC algorithm works as follows:

- *Step 1*: *(each IM)*: If the value of *PV(i, j)* counter is equal to or greater than *n*, send a request to the central arbiter.
- *Step 2*: *(central arbiter)*: If the central arbiter receives the request from *IM(i)*, it sets the value of the buffer load matrix element *x[i, j]* to 1 (the values of *i* and *j* come from the counter *PV(i, j)*).
- *Step 3*: *(central arbiter)*: After receiving all requests, the central arbiter tries to find an optimal set of 1s, which allows to send the most number of cells from IMs to OMs. The central arbiter has to go through all rows of the buffer load matrix to find a set of *n* 1s representing *IM(i)* and *OM(j)* matching. If there is not possible to find a set of *n* 1s it attempts to find a set of *(n-1)* 1s, and so on.
- *Step 4*: *(each IM)*: In the next time slot send *n* cells from IMs to the matched OMs. Decrease the value of *PV(i, j)* by *n*. For IM-OM pairs not matched by the central arbiter use the SD scheme and decrease the value of *PV* counters by 1.

The steps in the SD-FC scheme are the same as in the SD-OC scheme, but the optimization process in the third step is not carried out. The central arbiter chooses the first 1, which fulfill the requirements in each row. The row searched as the first one is selected according to the round robin routine.

4.3 Input-Output Module matching algorithm

The IOM packet dispatching scheme employs also the central arbiter to make a matching between each IM and OM. The cells are sent only between IM-OM pairs matched by the arbiter. The SD scheme is not used.

In detail, the IOM algorithm works as follows:

- *Step 1: (each IM)*: Sort the values of *PV(i, j)* in descending order. Send to the central arbiter a request containing a list of the OMs identifiers. The identifier of *OM(j)* to which *VOMQ(i, j)* stores the most number of cells should be placed on the list as the first one, and the identifier of *OM(s)* to which *VOMQ(i, s)* stores the least number of cells should be placed on the list as the last one.
- *Step 2: (central arbiter)*: The central arbiter analyzes one by one the requests received from IMs and checks if it is possible to match *IM(i)* with *OM(j)*, the identifier of which was sent as the first one on the list in the request. If the matching is not possible, because the *OM(j)* is matched with other IM, the arbiter selects the next OM on the list. The round-robin arbitration is employed for selection of *IM(i*) the request of which is analyzed as the first one.
- *Step 3: (central arbiter)*: The central arbiter sends to each IM confirmation with the identifier of *OM(t)*, to which the IM is allowed to send cells.
- *Step 4: (each IM)*: Match all output links *LI(i, r)* with cells from *VOMQ(i, t)*. If there is less than *n* cells to be sent to *OM(t)*, some output links remain unmatched.
- *Step 5: (each IM)*: Decrease the value of *PV(i, t)* by the number of cells which will be sent to *OM(t)*.
- *Step 6: (each IM)*: In the next time slot send the cells from the matched *VOMQ(i, t)* to the *OM(t)* selected by the central arbiter.

4.4 Performance of SD-FC, FD-OC and IOM schemes

The simulation experiments were carried out under the same conditions as the experiments for the distributed arbitration (see subchapter 3.6). We have evaluated two performance measures: average cell delay in time slots and maximum VOMQs size (we have investigated the worst case). The size of the buffers at the input and output side of switching fabric is not limited, so cells are not discarded. However, they encounter the delay instead. Because of the unlimited size of buffers, no mechanism controlling flow control between the IMs and OMs (to avoid buffer overflows) is implemented. The results of the simulation for the Bernoulli arrival model are shown in the charts (Fig. 23-32). Fig. 23, 25, 27, 29 show the average cell delay in time slots obtained for the uniform, Chang's, trans-diagonal, bi-diagonal, and bursty traffic patterns, whereas Fig. 24, 26, 28, 30 show the maximum VOMQ size in number of cells. Fig. 31, 32 show the results for the bursty traffic with the average burst size b=16, and uniform traffic distribution pattern.

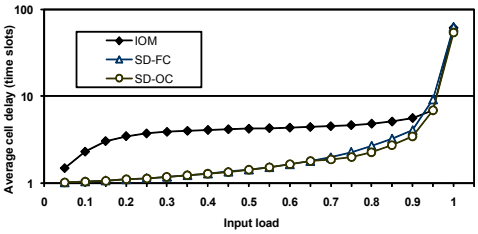


Fig. 23. Average cell delay, uniform traffic

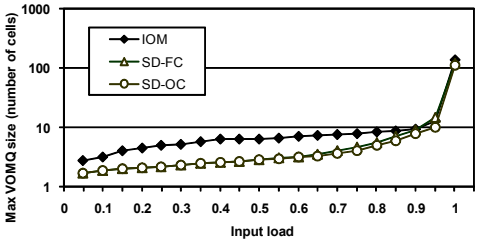


Fig. 26. The maximum VOMQ size, Chang's traffic

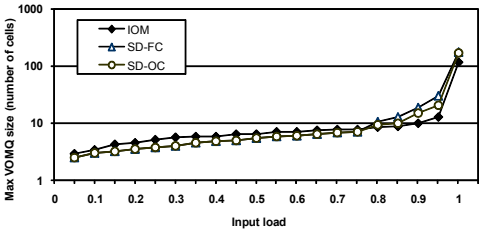


Fig. 24 The maximum VOMQ size, uniform traffic

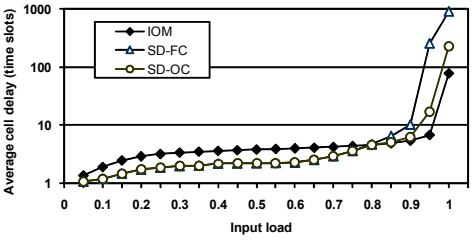


Fig. 27. Average cell delay, trans-diagonal traffic

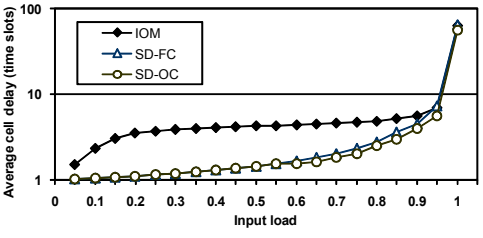


Fig. 25. Average cell delay, Chang's traffic

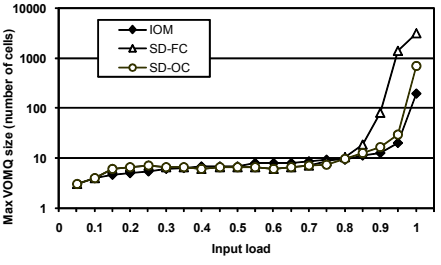


Fig. 28 The maximum VOMQ size, trans-diagonal traffic

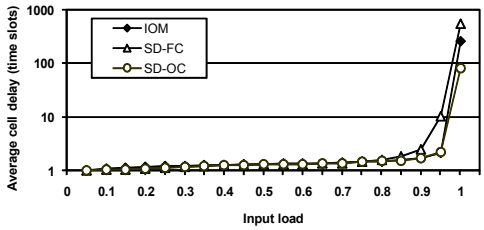


Fig. 29. Average cell delay, bi-diagonal traffic

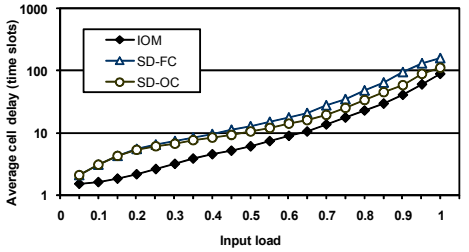


Fig. 31. Average cell delay, bursty traffic

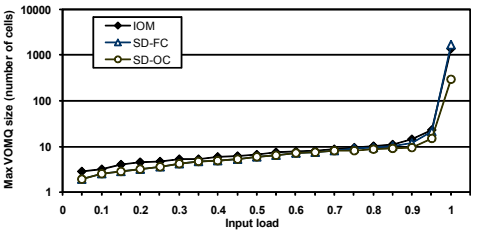


Fig. 30. The maximum VOMQ size, bi-diagonal traffic

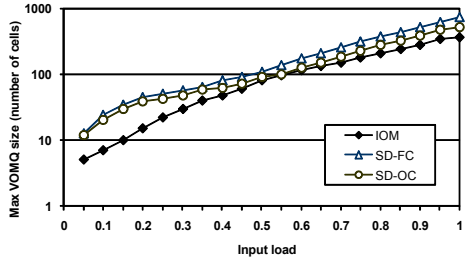


Fig. 32. The maximum VOMQ size, bursty traffic

We can see that the MSM Clos-network switch with all the schemes proposed achieves 100% throughput for all kinds of investigated traffic distribution patterns under Bernoulli arrival model and for the bursty traffic. The average cell delay is less than 10 for wide range of input load, regardless of the traffic distribution pattern. It is a very interesting result especially for the trans-diagonal and bi-diagonal traffic patterns. Both traffic patterns are highly demanding and many packet dispatching schemes proposed in the literature cannot provide the 100% throughput for the investigated switching fabric. For the bursty traffic, the average cell delay grows very similar to linear function of input load with the maximum value less than 150. We can see that the very complicated arbitration routine used in the SD-OC scheme does not improve the performance of the MSM Clos-network switch. In some cases the results are even worse than for IOM scheme (the trans-diagonal traffic with very high input load and the bursty traffic – Fig. 27 and 31). Generally, the IOM scheme gives higher latency than the SD schemes, especially for low to medium input load. It is due to matching *IM(i)* to that *OM(j)* to which it is possible to send the most number of cells. As a consequence, it is less probable to match IM-OM pairs to serve one or two cells per cycle.

The size of VOMQ in the MSM Clos switching network depends on the traffic distribution pattern. For all presented packet distribution schemes and the uniform and Chang's traffic the maximum size of VOMQ is less than 140 cells. It means that in the worst case, the average number of cell waiting for transmission to particular output was not bigger than 16. For the trans-diagonal traffic and the IOM scheme the maximum size of VOMQ is less than 200, but for the SD-OC and SD-FC the size is greater and come to 700 and 3000 respectively. For the bi-diagonal traffic the smallest size of VOMQ was obtained for the SD-OC scheme -

less than 290. For the bursty traffic the maximal size of VOMQ comes to: 750 for the SD-FC, 500 for the SD-OC and 350 for the IOM scheme.

5. Related Works

The field of packet scheduling in VOQ switches boasts of an extensive literature. Many algorithms are applicable to the single-stage (crossbar) switches and are not useful for packet dispatching in the MSM Clos-network switches. Some of them are more oriented to implementation, whereas others are of more theoretical significance. Here we review a representation of the works concerning packet dispatching in the MSM Clos-network switches.

Pipeline-Based Concurrent Round Robin Dispatching
E. Oki at al. have proposed in (Oki at al., 2002b) the Pipeline-Based Concurrent Round Robin Dispatching (PCRRD) scheme for the Clos-network switches. The algorithm can relax the strict timing constraint required by the CRRD and CMSD schemes. These algorithms have constrained dispatching scheduling to one cell slot. The constraint is a bottleneck when the switch capacity increases. The PCRRD scheme is able to relax the scheduling time into more than one time slot, however nk^2 request counters and P subschedulers have to be used to support the dispatching algorithm. Each subscheduler is allowed to take more than one time slot for packet scheduling, whereas one of them provides the dispatching result every time slot. The subschedulers adopt the CRRD algorithm, but other schemes (like CMSD) may be also adopted. Both, the centralized and non-centralized implementations of the algorithm are possible. In the centralized approach, each subscheduler is connected to all IMs. In the non-centralized approach, the subschedulers are implemented in different locations i.e. in IMs and CMs. The PCRRD algorithm provides 100% throughput under uniform traffic and ensures that cells from the same VOQ are transmitted in sequence.
Maximum Weight Matching Dispatching
The Maximum Weight Matching Dispatching scheme (MWMD) for the MSM Clos-network switches was proposed by R. Rojas-Cessa at al. in (Rojas-Cassa at al., 2004). The scheme is based on the maximum weight matching algorithm implemented in input-buffered single-stage switches. To perform the MWMD scheme each *IM(i)* has k virtual output-module queues (VOMQs) to eliminate HOL blocking. VOMQs are used instead of VOQs and *VOMQ(i, j)* stores cells at *IM(i)* destined to *OM(j)*. Each VOMQ is associated with m request queues (RQ), each denoted as *RQ(i, j, r)*. The request queue *RQ(i, j, r)* is located in *IM(i)* and stores requests of cells destined for *OM(j)* through *CM(r)* and keeps the waiting time *W(i, j,r)*. The waiting time represents the number of slots a head-of-line request has been waiting. When a cell enters *VOMQ(i, j)*, the request is randomly distributed and stored in *RQ(i, j, r)* among m request queues. A request in *RQ(i, j, r)* is not related to a specific cell but to *VOMQ(i, j)*. A cell is sent from *VOMQ(i, j)* to *OM(j)* in a FIFO manner when a request in *RQ(i, j, r)* is granted.
The MWMD scheme uses a central scheduler which consists of m subschedulers, denoted as *S(r)*. Each subscheduler is responsible for selecting requests related to cells which can be transmitted through *CM(r)* at the next time slot e.g.: subscheduler *S(0)* selects up to k requests from k^2 RQs, where corresponding cells to the selected RQs are transmitted through *CM(0)* at the next time slot. *S(r)* selects one request from each IM and one request to each OM according to the Oldest-Cell-First (OCF) algorithm. The OCF algorithm uses the waiting

time *W(i, j, r)* which is kept by each *RQ(i, j, r)* queue. *S(r)* finds a match *M(r)* at each time slot, so that the sum of *W(i, j, r)* for all *i* and *j*, and a particular *r* is maximized. It should be stressed that each subscheduler behaves independently and concurrently, and uses only k^2 *W(i, j, r)* to find *M(r)*.
When *RQ(i, j, r)* is granted by *S(r),* the HOL request in *RQ(i, j, r)* is dequeued and a cell from *VOMQ(i, j)* is sent at the next time slot. The cell is one of the HOL cells in *VOMQ(i, j)*. The number of cells sent to OMs is equal to the number of granted requests by all subschedulers. R. Cessa at al. has proved that the MWMD algorithm achieves 100% throughput for all admissible independent arrival processes without internal bandwidth expansion, i.e. *n=m* for the Clos MSM network.

Maximal Oldest Cell First Matching Dispatching
The Maximal Oldest-cell first Matching Dispatching (MOMD) scheme was proposed by R. Rojas-Cessa at al. in (Rojas-Cassa at al., 2004). The algorithm has lower complexity for a practical implementation than MWMD scheme. The MOMD scheme uses the same queues arrangement as MWMD scheme: *k* VOMQs at each IM, each denoted as *VOMQ(i, j)* and *m* request queues, RQs, each associated with a VOMQ, each denoted as *RQ(i, j, r)*. Each cell enters a *VOMQ(i, j)* gets a time stamp. A request with the time stamp is stored in *RQ(i, j, r)*, where *r* is randomly selected. The distribution of the requests can also be done in the round-robin fashion among RQs. The MOMD uses distributed arbiters in IMs and CMs. In each IM, there are *m* output-link arbiters, and in each CM there are *k* arbiters, each of which corresponds to a particular OM. To determine the matching between *VOMQ(i, j)* and the output link *LI(i, r)* each non-empty *RQ(i, j, r)* sends a request to the unmatched output link arbiter associated to *LI(i, r)*. The request includes the time stamp of the associated cell waiting at the HOL to be sent. Each output-link arbiter chooses one request by selecting the oldest time stamp, and sends the grant to the selected RQ and VOMQ. Then, each *LI(i, r)* sends the request to the *CM(r)* belonging to the selected VOMQ. Each round-robin arbiter associated with *OM(j)* grants one request with the oldest time stamp and sends the grant to *LI(i, r)* of *IM(i)*. If an IM receives a grant from a CM, the IM sends a HOL cell from that VOMQ at the next time slot. There is possible to consider more iteration between IM and CM within the time slot.
The delay and throughput performance of 64×64 Clos-network switch, where *n=m=k=8* under MOMD scheme are presented in (Rojas-Cassa at al., 2004). The scheme cannot achieve the 100% throughput under uniform traffic with a single IM-CM iteration. The simulation shows that CRRD scheme is more effective under uniform traffic than the MOMD, as the CRRD achieves high throughput with one iteration. However, as the number of IM-CM iterations increases, the MOMD scheme gets higher throughput e.g. in the switch under simulation, the number of iterations to provide 100% throughput is four. The MOMD scheme can provide high throughput under a nonuniform traffic pattern (opposite to the CRRD scheme), called unbalanced, but the number of IM-CM iterations has to be increased to eight. The unbalanced traffic pattern has one fraction of traffic with uniform distribution and the other faction *w* of traffic destined to the output with the same index number as the input; when *w=0*, the traffic is uniform; when *w=1* the traffic is totally directional.

Frame Occupancy-Based Random Dispatching and Frame Occupancy-Based Concurrent Round-Robin Dispatching

The Frame occupancy-based Random Dispatching (FRD) and Frame occupancy-based Concurrent Round-Robin Dispatching (FCRRD) schemes were proposed by C-B. Lin and R. Rojas-Cessa in (Lin & Rojas-Cessa, 2005). Frame based scheduling with fixed-size frames was first introduced to improve switching performance in one-stage input-queued switches. C-B. Lin and R. Rojas-Cessa adopted captured-frame concept for the MSM Clos-network switches using RD and CRRD schemes as the basic dispatching algorithms. The frame concept is related to a VOQ and means the set of one or more cells in a VOQ that are eligible for dispatching. Only the HOL cell of the VOQ is eligible per time slot. The captured fame size is equal to the cell occupancy at *VOQ(i, j, l)* at the time t_c of matching the last cell of the frame associated to *VOQ(i, j, l)*. Cells arriving to *VOQ(i, j, l)* at time t_d, where $t_d > t_c$, are considered for matching if a new frame is captured. Each VOQ has a captured-frame size counter denoted as $CF_{i,j,l}(t)$. The value of this counter indicates the frame size at time slot *t*. The $CF_{i,j,l}(t)$ counter takes a new value when the last cell of the current frame of *VOQ(i, j, l)* is matched. Within the FCRRD scheme the arbitration process includes two phases and the request-grant-accept approach is implemented. The achieved match is kept during the frame duration.

The FRD and FCRRD schemes show higher performance under uniform and several nonuniform traffic patterns, as compared to the RD and CRRD algorithms. What's more the FCRRD scheme with two iterations is sufficient to achieve a high switching performance. The hardware and timing complexity of the FCRRD is comparable to that of the CRRD.

Maximal Matching Static Desynchronization Algorithm

The Maximal Matching Static Desynchronization algorithm (MMSD) was proposed by J. Kleban and H. Santos in (Kleban & Santos, 2007). The MMSD scheme uses the distributed arbitration with the request-grant-accept handshaking approach but minimizes the number of iterations to one. The key idea of the MMSD scheme is static desynchronization of arbitration pointers. To avoid collisions in the second stage, all IMs use connection patterns that are static but different in each IM; it forces cells destined to the same OM, but located in different IMs, to be sent through other CMs. In the MMSD scheme two phases are considered for dispatching from the first to the second stage. In the first phase each IM selects up to m VOMQs and assigns them to IM output links. In the second phase requests associated with output links are sent from IM to CM. The arbitration results are sent from CMs to IMs, so the matching between IMs and CMs can be completed. If there is more than one request for the same output link in a CM, a request is granted from this IM which should use a given CM for connection to an appropriate OM, according to the fixed IM connection pattern. If requests come from other IMs, CM grants one request randomly. In each *IM(i)* there is one group pointer *PG(i, h)* and one *PV(i, v)* pointer, where $0 \le v \le nk - 1$. In *CM(r)*, there are *k* round robin arbiters, and each of them corresponds to *LC(r, j)* – an output link to the *OM(j)* – and has its own pointer *PC(r, j)*.

The performance results obtained for the MMSD algorithm are better or comparable with results obtained for other algorithms, but the scheme is less hardware-demanding and seems to be implementable with the current technology in the three-stage Clos-network switches.

The modified MSM Clos switching fabric with SDRUB packet dispatching scheme
The modified MSM Clos switching fabric and a very simple packet dispatching scheme, called Static Dispatching with Rapid Unload of Buffers (SDRUB) were proposed by J. Kleban at al. in (Kleban at al., 2007). The main idea of modification of the MSM Clos switching fabric lies in connecting bufferless CMs to the two-stage buffered switching fabric so as to give the possibility of rapid unload of VOMQs. In this way an expansion in IMs and OMs is used. The maximum number of connected CMs is equal to *m-1*, but it is possible to use less CMs. In practice, the number of CMs significantly influences the performance of the switching fabric. The number of CMs depends on the traffic distribution pattern to be served. Contrary to the MSM Clos switching fabric, in the modified architecture, at each time slot, it is possible to send one cell from each IM to each OM due to direct connecting path between IMs and OMs. The arbitration is necessary for rapid unload of buffers only.
In the SDRUB scheme each VOMQ has its own counter *PV(i, r)* which shows the number of cells destined to *OM(r)*. The SDRUB algorithm uses a central arbiter to indicate the IMs which are allowed to send cells through CMs. Assume that there is *(y-1)* CMs in the modified MSM Clos switching fabric. When *PV(i, r)* reaches the value equal or greater than *y*, it sends the information about the overloaded buffer to the central arbiter. In the central arbiter there is a binary matrix of buffers load. If the value of matrix element *x[i, j]* is 1, it means that *IM(i)* can send *y* cells to *OM(j)*, one through the direct connection and *y-1* through CMs. The central arbiter changes the value of element *x[i, j]* from 0 to 1 only if it is the first 1 in column *i* and row *j*. In other cases the request is rejected. The OM to which *IM(i)* sends cells using CMs is selected according to the round robin routine. No other optimization process for selecting IM-OM pairs for buffers rapid unload is employed.
Simulation experiments have shown that the modified MSM Clos switching fabric achieves very good performance under uniform as well as nonuniform traffic distribution patterns. To manage the trans-diagonal traffic effectively, it is necessary to implement at least *n/2* CMs. For such number of CMs the switching fabric achieves 100% throughput but any smaller number of CMs reduces the throughput of the switching fabric. Under the bi-diagonal traffic the SDRUB algorithm can achieve 100% throughput only when the maximum number of CMs is used. It is obvious that when the number of CMs increases, the throughput increases proportionally. For the uniform traffic pattern the SDRUB scheme gives very good results for one CM.

6. References

Chao, H. J., & Liu, B. (2007). *High Performance Switches and Routers*, John Wiley & Sons, Inc., ISBN: 978-0-470-05367-6, New Jersey

Chao, H. J., Cheuk, H. L. & Oki, E. (2001). *Broadband Packet Switching Technologies: A Practical Guide to ATM Switches and IP Routers,* John Wiley & Sons, Inc., ISBN: 0-471-00454-5, New York

Clos, C. (1953). A Study of Non-Blocking Switching Networks, *Bell Sys. Tech. Jour.*, Vol. 32, pp. 406-424

Hui, J. Y. & Arthurs, E. (1987). A Broadband Packet Switch for Integrated Transport, *IEEE J. Sel. Areas Commun.*, Vol. 5, No. 8, pp. 1264-1273

Jiang, Y. & Hamdi, M. (2001). A fully desynchronized round-robin matching scheduler for a VOQ packet switch architecture", *Proceedings of IEEE High Performance Switching and Routing 2001 – HPSR 2001,* pp. 407–411, US, Texas, Irving

Kabacinski, W. (2005). *Nonblocking Electronic and Photonic Switching Fabrics*, Springer, ISBN: 978-0-387-25431-9

Kleban, J. & Santos, H. (2007). Packet Dispatching Algorithms with the Static Connection Patterns Scheme for Three-Stage Buffered Clos-Network Switches, P*roceedings of IEEE International Conference on Communications 2007 - ICC-2007*, Scotland, Glasgow

Kleban, J. & Wieczorek, A. (2006). CRRD-OG - A Packet Dispatching Algorithm with Open Grants for Three-Stage Buffered Clos-Network Switches, P*roceedings of High Performance Switching and Routing 2006 – HPSR 2006*, pp. 315-320, Poland, Poznan

Kleban, J., Sobieraj, M. & Węclewski, S. (2007). The Modified MSM Clos Switching Fabric with Efficient Packet Dispatching Scheme", *Proceedings of IEEE High Performance Switching and Routing 2007 – HPSR 2007*, US, New York

Lin, C-B & Rojas-Cessa, R. (2005). Frame Occupancy-Based Dispatching Schemes for Buffered Three-stage Clos-Network switches", P*roceedings of 13th IEEE International Conference on Networks 2005,* Vol. 2, pp. 771-775.

McKeown, N., Mekkittikul, A., Anantharam, V. & Walrand, J. (1999), Achieving 100% Throughput in an Input-queued Switch, *IEEE Trans. Commun.,* Vol. 47, Issue 8, pp. 1260-1267

Oki, E., Jing, Z. & Rojas-Cessa, R. & Chao H. J. (2002a). Concurrent Round-Robin-Based Dispatching Schemes for Clos-Network Switches, *IEEE/ACM Trans. on Networking*, Vol. 10, No.6, pp. 830-844

Oki, E., Rojas-Cessa, R. & Chao, H. J. (2002b). PCRRD: A Pipeline-Based Concurrent Round-Robin Dispatching Scheme for Clos-Network Switches, P*roceedings of IEEE International Conference on Communications 2002 - ICC-2002*, pp. 2121-2125, US, New York

Pun, K., & Hamdi, M. (2004). Dispatching schemes for Clos-network switches, *Computer Networks* No. 44, pp.667-679

Rojas-Cessa, R. Oki, E. & Chao, H. J. (2004). Maximum Weight Matching Dispatching Scheme in Buffered Clos-Network Packet Switches, *Proceedings of IEEE International Conference on Communications 2004 - ICC-2004,* pp. 1075-1079, France, Paris

Yoshigoe, K. & Christensen, K. J. (2003). An evolution to crossbar switches with virtual ouptut queuing and buffered cross points, *IEEE Network*, Vol. 17, No. 5, pp. 48-56

8

RAS Modeling of a Large InfiniBand Switch System

Dong Tang and Ola Torudbakken
Sun Microsystems, Inc.
USA

1. Introduction

Computer clusters or grids constructed from open and standard commercial off the shelf (COTS) systems now dominate the top 500 supercomputer sites (Top500, 2008), providing an attractive way to rapidly construct high performance computing (HPC) systems of interconnected nodes. The largest of these HPC systems are now driving toward petascale deployments, delivering petaflops of computational capacity and petabytes of storage capacity. However, designing and building these large HPC systems involves significant challenges, including:

- Rapidly building and expanding the computational capacity of HPC clusters to meet growing demands
- Increasing levels of computational density while staying within constrained envelopes of power and cooling
- Reducing complexity and cost for physical infrastructure and management
- Implementing interconnect technology that can connect hundreds or thousands of processors without introducing unacceptable levels of latency

Interconnect technology plays a vital role in addressing all of these issues. InfiniBand has emerged as a compelling interconnect technology, and now provides more scalability and significantly better cost-performance than any other known fabric. In spite of its ability to provide high-speed connectivity and low latency, connecting and cabling thousands of compute nodes with smaller discrete InfiniBand switches remains problematic. With traditional approaches, the largest HPC clusters can require hundreds of switches, as well as thousands of ports and cables for inter-switch connectivity alone. The result can be significant added cost and complexity, not to mention energy and space consumption.

To address these challenges, the Sun Datacenter Switch 3456 (DS3456) system (Sun Microsystems, 2007) provides the world's largest standards-based DDR (dual data rate) InfiniBand switch, with direct capacity to host up to 3,456 server nodes. Only slightly larger than two conventional datacenter racks, the system drastically reduces the cost, power, and footprint of deploying very large-scale standards-based high performance computing

fabrics. DS3456 is tightly integrated with the Sun Blade 6048 modular rack system (Sun Microsystems, 2008) which supports InfiniBand leaf switch, facilitating deployment of HPC systems up to 13,824 Nodes. Together these technologies offer low latency, high compute density, reduced cabling and management complexity, and lower power consumption than with other solutions.

Given this new large switch system, an important issue that needs to be addressed is the quantification of the associated RAS features. In this study, we developed a hierarchical Markov availability model (Trivedi, 2001) for DS3456 to assess its reliability, availability, and serviceability (RAS), using RAScad (Tang et al., 2002), a Sun internal RAS modeling tool that supports hierarchical modeling and automatic model generation.

The rest of this chapter is organized as follows: Section 2 gives an overview of Sun DS3456; Section 3 defines RAS metrics; Section 4 describes the model and parameters; Section 5 presents results and analysis; and Section 6 concludes the study.

2. Overview of DS3456

InfiniBand is a technology developed to address low-latency, high-performance, and low overhead communications between servers and I/O devices. It defines an architecture of networking principles – switching and routing – to provide a scalable, high-performance server I/O fabric (Cisco Systems, 2006). InfiniBand is a loss-less interconnect providing ordered packet delivery across the fabric through the use of credit-based flow-control. To ensure data integrity, its end-to-end protocols include fault tolerant features such as link-level and end-to-end CRC, packet re-transmission, multi-path routing, and automatic path migration. Upper-layer protocols, built on top of these provisions, allow seamless fit into existing networking and storage protocols. In addition, QoS (Quality of Service) and congestion control mechanisms are natively included in InfiniBand. All of these provide an excellent, converged fabrics solution for running storage, networking and clustering traffic.

DS3456 is the world's largest InfiniBand switch system, with capacity for connection of up to 3,456 nodes. The basic switch element used in DS3456 is the InfiniScale III (IS3) 24-port InfiniBand switch chip (Mellanox Technologies, 2009). The DDR version of IS3 supports 16 Gbps per 4x port, delivering up to 768 Gbps of aggregate bandwidth. The chip architecture features an intelligent non-blocking packet switch design with an advanced scheduling engine that provides QoS with switching latencies of less than 140 nanoseconds. DS3456 has been deployed in several HPC systems, including Ranger, the world No. 6 HPC system with peak performance of 579.4TFlops (Top500, 2008), located at Texas Advanced Computing Center, University of Texas at Austin.

Figure 1 is the physical view of DS3456. The major high-level DS3456 components and related RAS features are listed as follows.

- Twenty-four horizontally-installed line cards with each providing 48 12x connectors delivering 144 DDR 4x InfiniBand ports. Each line card connects to pass-through connectors in a passive orthogonal midplane.
- Eighteen vertically-installed fabric cards directly connected to the line cards through the orthogonal midplane. Each fabric card also features eight modular high-

performance fans that provide front-to-back cooling for the chassis. The eight fans are N+1 redundant and hot swappable.

- Two fully-redundant chassis management controller cards (CMCs) monitoring all critical chassis functions including power, cooling, line cards, fabric cards, and fan modules. CMC is hot swappable.
- Sixteen power supply units (PSUs) divided into two banks of eight units, with each bank providing N+1 redundant PSUs to half the line cards and half the fabric cards. PSU is hot swappable.

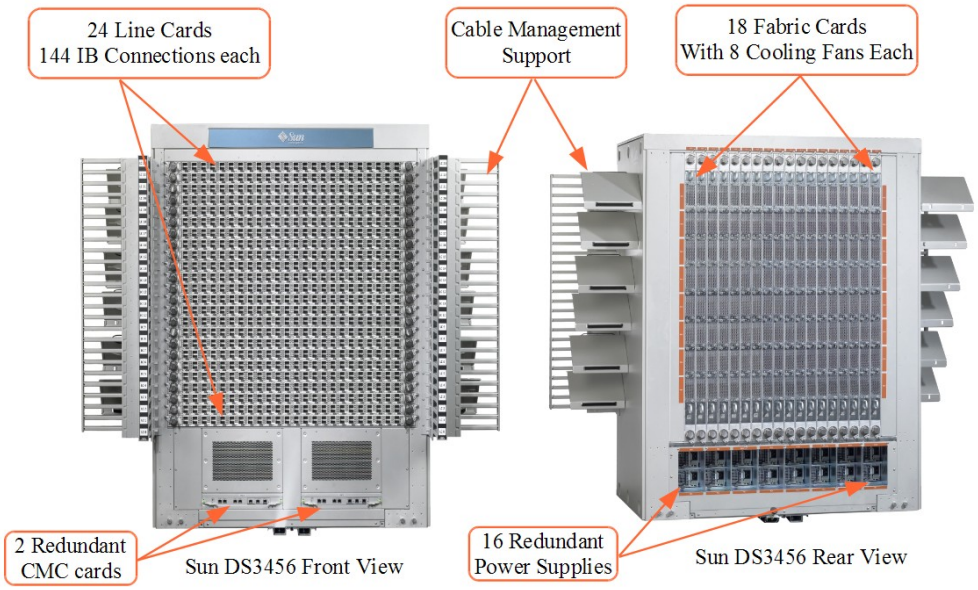


Fig. 1. DS3456 Physical View

Figure 2 shows the connectivity between line cards and fabric cards for DS3456. The passive midplane provides 432 8x8 orthogonal connectors arrayed in an 18x24 grid. Each line card contains 24 IS3 switch chips, 12 interfacing to the midplane, and 12 interfacing to the 12x connectors at the front of line card. A total of 144 4x InfiniBand ports are provided by each line card, expressed as 48 physical 12x connectors. Each fabric card contains eight IS3 switch chips connected to the midplane, providing interconnect between different line cards.

Thus, a communication path starts from an external port connected to an IS3 chip at the bottom row of a line card, goes through an IS3 chip at the top row of the same line card, an IS3 chip on a fabric card, two IS3 chips on the destination line card (one at the top row and one at the bottom row), and ends at another external port connected to the destination IS3 chip. That is, a message packet goes through as many as five stages of switch from the source port to the destination port.

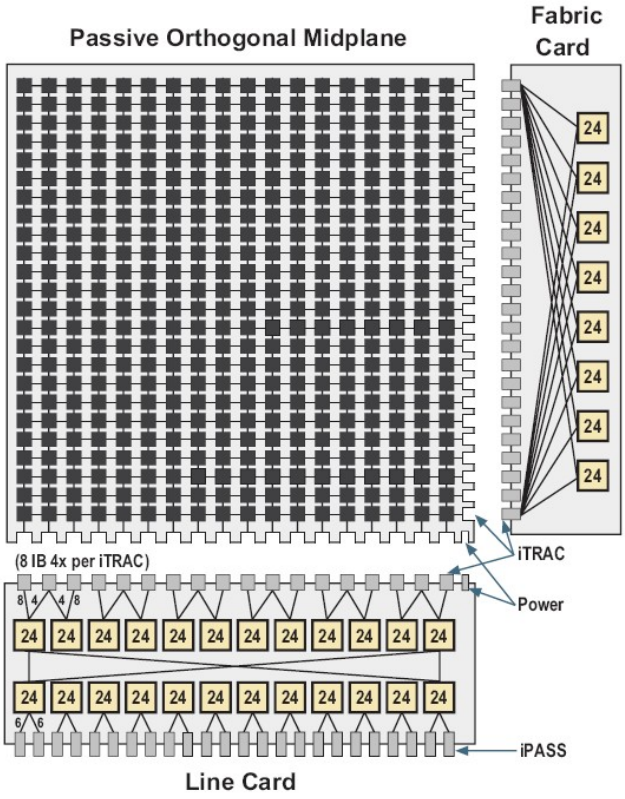


Fig. 2. DS3456 Internal Connectivity

3. RAS metrics defined

To quantify RAS for the target system, we first define RAS metrics and related concepts in this section. For simplicity, the capacity of the target system is assumed to be fully used, i.e., all 3,456 ports of the switch are utilized to connect server nodes.

3.1 Reliability

Connectivity between the server nodes using the switch for communication is a reliability measure for the switch system. A connectivity failure is defined as the loss of communication between a server node physically connected to the switch and another server node physically connected to the same switch due to hardware problems in the switch. We use *Mean Time Between Connectivity Failures* (MTBCF) to quantify reliability for the switch system.

A line card or fabric card failure would cause partial communication paths in the switch to be unavailable. Unavailability of partial paths caused by a fabric card failure does not affect connectivity as paths that were routed across the faulty fabric card can be re-routed to the operational fabric cards. Unavailability of partial paths caused by a line card failure may or

may not translate to connectivity failures, depending on redundancy in the interconnect topology between the switch and server nodes.

- Non-redundancy case. If each server node connects to only one port on the switch, a line card failure would result in connectivity failures for some of the server nodes connected to the switch.
- Redundancy case. Typically, each server node connects to two or four ports on different line cards in the switch. In this case, unavailability of partial paths caused by the failure of one line card does not generate any connectivity failures.

3.2 Availability

The traditional availability definition is the proportion of time that the system is operational and delivering required services. At any time point, the system is in either an up or a down state. However, for a degradable system, the system can be in a partially available state. For the non-redundancy case of DS3456, unavailability of partial paths does not disable the function of the entire switch, but degrades system capacity. Thus, the system can be in partially available states, in addition to the fully available and failure states. For instance, when a line card fails, paths related to the faulty line card (out of 24 line cards) are unavailable and the system capacity is reduced by 1/24. Therefore, we defined availability for this state as 23/24. The RAScad performability (Trivedi, 2001) evaluation capability is used to generate this performance-oriented availability.

3.3 Service cost

In traditional service strategies, every component failure in the system translates to a service call. For such a large system as DS3456, replacing a line card or fabric card is particularly time consuming, because it may take several hours for the system to complete the restart process after a power-off repair. It is thus desirable to reduce service frequency as much as possible.

Previous studies showed that adoption of deferred repair service strategies for redundant components can greatly reduce unscheduled service events and associated system downtime (Sun, 2005). In this study, we once again analyzed the effect of deferred repair on system availability and service cost for the redundancy case. We use *Unscheduled Mean Time Between Services* (U_MTBS) to quantify service cost for the switch system.

3.4 Failure rate estimation

These metrics are calculated from a system-level RAS model built by utilizing information on the system configuration and its RAS characteristics (redundancy, hot or cold swap, etc.), applying a failure rate to each component, and then integrating them into the model. These failure rates are estimated from previous field data using the field-based *Mean Time Between Failures* (MTBF) prediction method described below where MTBF = 1/failure rtae.

Field-Based MTBF Prediction Method — The Field Replaceable Unit (FRU) MTBFs are calculated using methods described in Telcordia TR-NWT-000332 (Telcordia Technologies, 2001) with lower component-level (ICs, resistors, capacitors, etc.) failure rates adjusted based on field data, or directly estimated from field data, or provided by the OEM vendors.

The field data used to calibrate component failure rates were collected from tens of thousands of Sun field systems with billions of cumulative operating hours. This approach is called the Sun field-based MTBF prediction method.

4. RAS model and parameters

Similar to many studies of this type, we assumed independent failures on different components and constant failure rates. The target system is modeled as a hierarchy of Markov chains. The top level model is shown in Figure 3. In a RAScad Markov model, the user can define three reward vectors for each state, as displayed in the circles representing states (Tang & Trivedi, 2004): (1) Availability (0 or 1), (2) Performance (≥ 0), and (3) Service Cost (≥ 0).

The first reward vector is used to calculate system availability. The second reward vector is used to calculate system performability. The third reward vector is used to calculate annual service cost or service call rate. In the DS3456 model, up to two failures of line card and fabric card, which have impact on system performability (for the non-redundancy case), were modeled in detail. The notation used in the models is explained as follow:

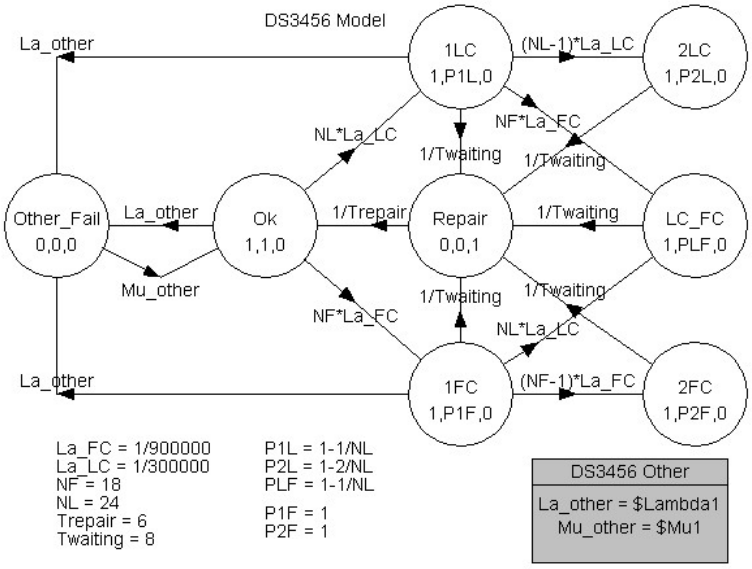


Fig. 3. Top level Markov model

- Ok: state in which the system is functioning properly (no faults)
- 1LC: state in which one line card has failed
- 2LC: state in which two line cards have failed
- 1FC: state in which one fabric card has failed
- 2FC: state in which two fabric cards have failed
- LC_FC: state in which one line card and one fabric card have failed
- Repair: state in which the system is shutdown to replace faulty line card or fabric card

- Other_Fail: state in which the system is down due to other hardware component failures
- NL: number of line cards in the system (18)
- NF: number of fabric cards in the system (24)
- Twaiting: service waiting time – waiting for off-peak hours to repair the system (8 hours)
- Trepair: repair time including restart time (6 hours)
- La_LC: failure rate for line card (1/900K hours)
- La_FC: failure rate for fabric card (1/300K hours)
- La_other: system failure rate due to other hardware faults (calculated from submodel)
- Mu_other: system repair rate for other hardware faults (calculated from submodel)

When one or more line/fabric cards have failed (states 1LC, 1FC, 2LC, 2FC, and LC_FC), the system is scheduled to be shutdown for repair after a service waiting time. For the non-redundancy case, these states may be degraded states, as shown by the performance reward vector in these states (P1L, P2L, etc.). For the redundancy case, these states are still fully functioning states, as shown by the availability reward vector in the states (all values are 1).

In Figure 3, the gray color rectangle box represents the interface between the current model and the submodel called DS3456 Other. All hardware components other than line cards and fabric cards are included in the submodel (details are not discussed in this chapter). If a system failure occurs due to hardware problems other than line card and fabric card faults, the system goes from the Ok state to the Other_Fail state. The associated failure rate (La_other) and repair rate (Mu_other) are bound to the submodel output Lambda1 and Mu1 which are the equivalent failure rate and repair rate (Lanus et al., 2003) of the submodel.

The model parameters, as listed above, were estimated using the Sun field-based MTBF prediction method discussed in Section 3.4 or based on engineering judgements. The repair time was estimated to be 6 hours because the system restart time is long.

5. Analysis of results

In this section, we present RAS results for the target system, including basic results, interval results (assuming deferred repair), and uncertainty analysis on key parameters.

5.1 Basic results

Table 1 shows the steady-state system level results evaluated from the DS3456 model by RAScad. The results show that for the redundancy case, MTBCF is much longer than that for the non-redundancy case. That is, with two or four redundant ports on different line cards, the system reliability is high in terms of connectivity. But this is not the case for system availability due to the large number of line/fabric cards and the long duration of power-off repair time of these cards. The system availability is similar for both redundancy and non-redundancy cases. This is because the system unavailability is dominated by power-off repair events, which are common for both cases. In other words, the system unavailability is not significantly affected by the degraded states for the non-redundancy case.

Configuration	U_MTBS (hours)	MTBCF (hours)	Availability
Non-redundancy	5,937	9,679	0.999372
Redundancy	5,937	3.23E6	0.999398

Table 1. Steady-state results for DS3456

A high availability DS3456 configuration typically implements interconnect between a server node and two (2-way redundancy) or four (4-way redundancy) different line cards, utilizing standard 4x InfiniBand ports. In the following, our discussion is focused on the 4-way redundancy configuration. To investigate which components in the system contribute most to the system unavailability (or downtime) and service events, we did a breakdown analysis as shown in Figures 4 and 5.

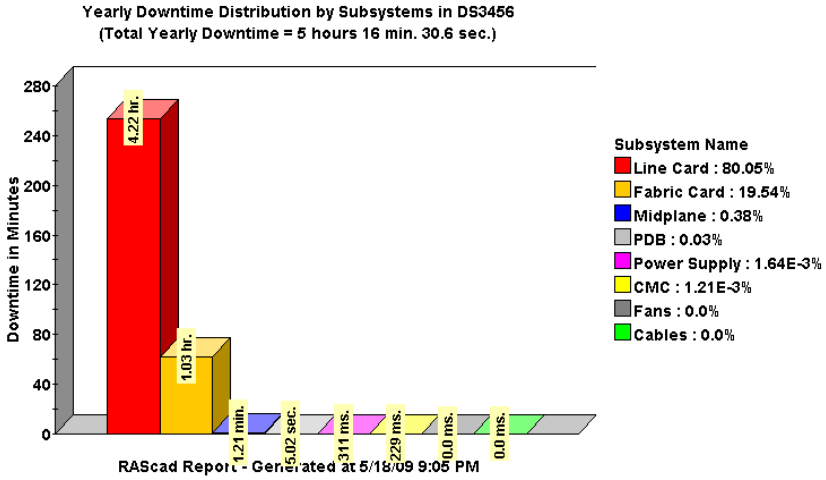


Fig. 4. Distribution of system downtime

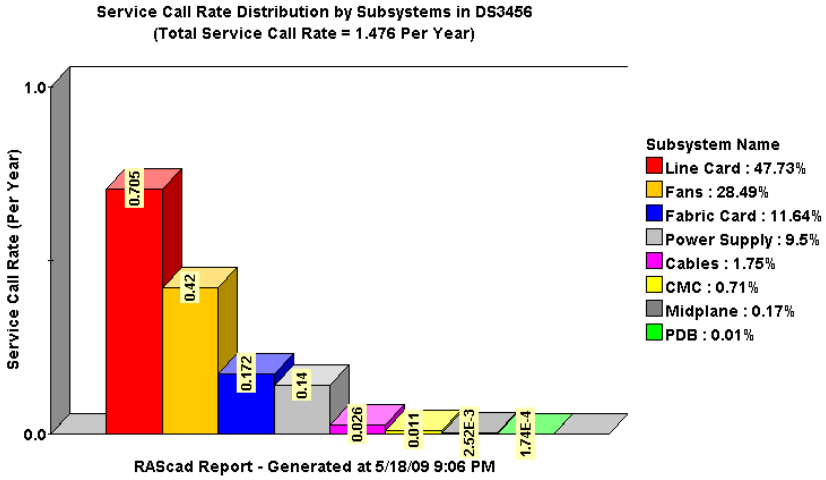


Fig. 5. Distribution of service events

Figure 4 shows that the system unavailability is dominated by shutdown repairs for faulty line cards and fabric cards. Figure 5 shows that the service events are mostly due to the

following components: line cards, fans, fabric cards, and power supply units. Deferred repair of these components, if possible, could significantly reduce unscheduled service events and system downtimes. For the 4-way redundancy configuration, we can tolerate at least two line card or fabric card failures without losing any connectivity. Since every eight fans (N+1 redundant) are associated with a fabric card, we can also tolerate the failure of two fans associated with a line card (equivalent to a line card failure) or three fans otherwise.

5.2 Deferred repair

Given these thresholds of component failures that can be tolerated without degrading system performance, the following deferred repair service strategy is proposed for the target system. The system is serviced periodically, referenced as scheduled service, according to a predefined maintenance schedule, to repair all the components that have failed since the last service event. During the time window between two scheduled services, an unscheduled service is triggered upon any of the following events:

- Two line cards have failed.
- Two fabric cards have failed.
- One line card and one fabric card have failed.
- Two fans associated with a fabric card or any three fans have failed.
- Any other hardware component failures that stop the functioning of system (e.g., failure of two PSUs in a power bank).

The Markov model in Figure 3 can be easily modified to model this deferred repair service strategy by removing the transition from state 1LC to state Repair and the transition from state 1FC to state Repair. That is, no repair action is taken upon a failure of line card or fabric card. In addition, one of the submodels in the hierarchy, the fan model, also needs to be modified, as shown in Figure 6. In the diagram, La_fan is the fan failure rate and N is the total number of fans in the system. The failure of two fans associated with a fabric card is modeled by the transition from state 1Fan_Down to state Repair. The failure of any three fans is modeled by the transition from state 2Fan_Down to state Repair.

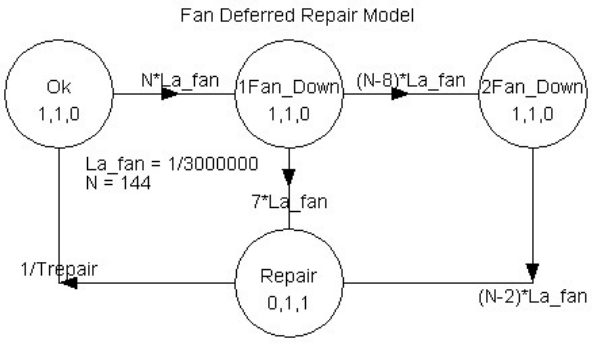


Fig. 6. Deferred repair model for fans

Table 2 shows the interval system-level results for different service strategies generated by RAScad. Our previous study (Tang & Trivedi, 2004) showed that the interval availability (average availability for a time interval from 0 to T) and associated measures, such as interval failure rate and interval service call rate, instead of steady-state measures, should be used for systems with deferred repair. In the table, "Deferred, 3 months" means a deferred repair service strategy with a periodic maintenance schedule of three months.

Service Strategy	U_MTBS (hours)	Unscheduled Yearly Downtime
No deferred repair	5,937 hr.	5 hr. 16.25 min.
Deferred, 1 month	41,992 hr.	11.44 min.
Deferred, 3 months	33,527 hr.	29.98 min.
Deferred, 6 months	26,941 hr.	52.04 min.

Table 2. Interval results for DS3456 under different service strategies

The table indicates that adoption of the proposed deferred repair service strategies can significantly reduce unscheduled service events as well as system downtime. With a quarterly maintenance schedule, the unscheduled MTBS is five times longer and the unscheduled system downtime is reduced by 90%. With a monthly maintenance schedule, the unscheduled MTBS is seven times longer and the unscheduled system downtime is reduced by 96%.

Although the system reliability and unscheduled downtime can be further improved by increasing maintenance frequency, the scheduled downtime (6 hours for each maintenance event) will also increase, leading to lower overall system availability. Given a tradeoff between system reliability and availability, we recommend deferred repair service strategies with a maintenance time window of 1 to 3 months.

5.3 Uncertainty analysis

Two key parameters in the model are the line card and fabric card failure rates. How sensitive are the results to the variance of these parameters? To answer this question, we performed an uncertainty analysis using RAScad. In each experiment of the analysis, the two parameters were randomly selected from the ±50% range of the estimated mean value, respectively, to generate a point of result. The sample size is 1,000. Figure 7 and Figure 8 plot the results for the "Deferred, 3 months" service strategy.

Figure 7 shows that for unscheduled MTBS, the 90% confidence interval is (26856, 40979), with the mean of 33,490 hours. That is, U_MTBS is likely to vary about ±20% around the mean when the uncertainty of the two key parameters is ±50%. Figure 8 shows that for unscheduled yearly downtime, the 90% confidence interval is (12.8, 52.7), with the mean of 30.5 minutes. That is, the system availability is most likely to stay at the 0.9999 level (equivalent to 5.3 to 53 minutes of yearly downtime), given the ±50% uncertainty of the two key parameters. Notice the slight difference between these means estimated from the simulations and those calculated numerically in Table 2. This is due to the nature of random sampling in simulations.

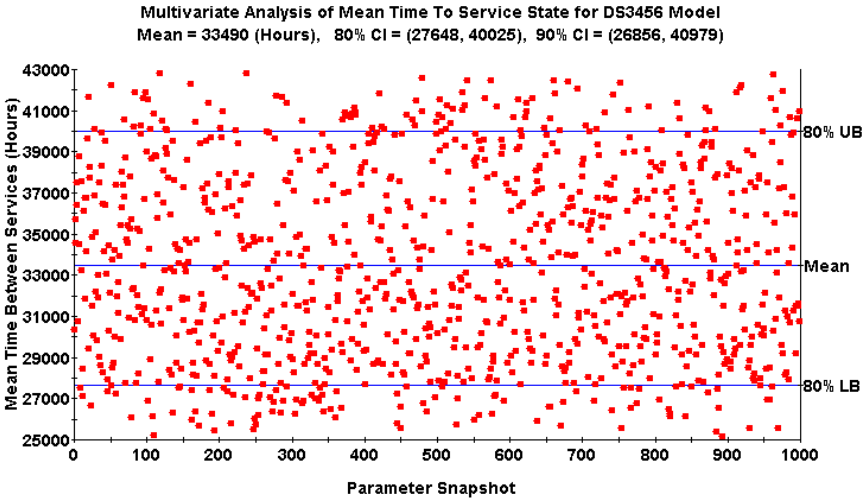


Fig. 7. Uncertainty analysis plot for U_MTBS

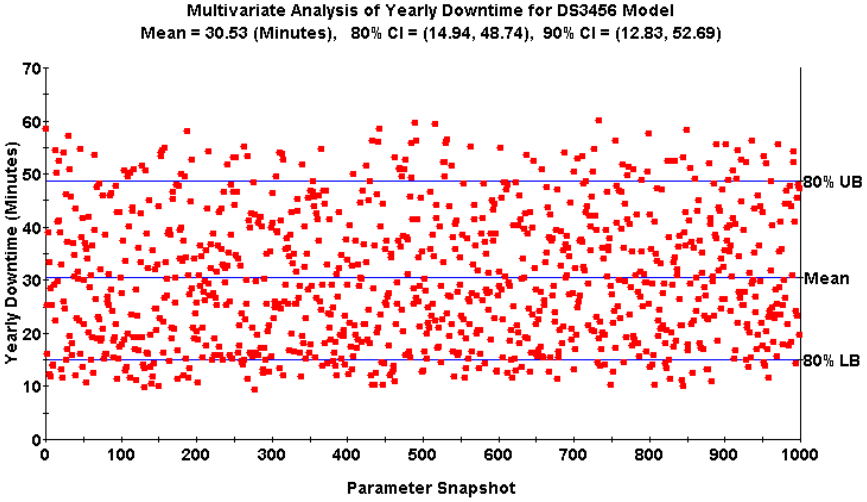


Fig. 8. Uncertainty analysis plot for yearly downtime

6. Conclusion

In this chapter, we presented a reliability, availability, and serviceability modeling and analysis, against hardware faults, for the Sun Datacenter Switch 3456 system, the world's largest InfiniBand switch system. To our knowledge, this is the first effort in RAS modeling of such a large switch system. The study demonstrated how the hierarchical Markov

modeling approach can be used on large switch systems to reduce model complexity and therefore the feasibility of RAS quantification for large switch systems.

The results show that the system reliability, in terms of connectivity between the server nodes physically connected to the switch, is high for configurations with redundant ports (MTBCF > 3 million hours). The study also investigated the RAS benefits of practicing deferred repair service strategies and identified optimal maintenance time windows. Adoption of our recommended service strategies can significantly reduce unscheduled service events (five to seven times longer U_MTBS) and system downtime (by 90% to 96%). Finally, an uncertainty analysis was performed to study the sensitivity of results to the variance of key parameters. The analysis generated 90% confidence intervals of system RAS measures for the ±50% uncertainty of two key parameters.

7. References

Cisco Systems (2006). *Understanding Infiniband*, White Paper, http://www.cisco.com/

Lanus M.; Ying L. & Trivedi K. S. (2003). Hierarchical Composition and Aggregation of State-based Availability and Performability Models, *IEEE Transactions on Reliability*, Vol. 52, No. 1, March 2003, pp. 44-52

Mellanox Technologies (2009). *InfiniScale III Product Brief*, http://www.mellanox.com/

Sun H.; Tang D. & Wood R. (2005). Optimizing Service Strategy for Systems with Deferred Repair, *Proceedings of the 11th Pacific Rim International Symposium on Dependable Computing* (PRDC'05), ISBN 0-7695-2492-3, Changsha, China, Dec. 2005, IEEE, Los Alamitos, California

Sun Microsystems (2007). *Sun Datacenter 3456 Switch System Architecture*, White Paper, Nov. 2007.

Sun Microsystems (2008). *Pathways to Petascale Computing: The Sun Constellation System – Designed for Performance*, White Paper, Feb. 2008.

Tang D.; Zhu J. & Andrada R. (2002). Automatic Generation of Availability Models in RAScad, *Proceedings of International Conference on Dependable Systems and Networks*, (DSN 2002), pp. 488-492, ISBN 0-7695-1597-5, Washington DC, USA, June 2002, IEEE, Los Alamitos, California

Tang D. & Trivedi K. S. (2004). Hierarchical Computation of Interval Availability and Related Metrics, *Proceedings of International Conference on Dependable Systems and Networks* (DSN 2004), pp. 693-698, ISBN 0-7695-2052-9, Florence, Italy, June 2004, IEEE, Los Alamitos, California

Telcordia Technologies (2001). *SR332 - Reliability Prediction Procedure of Electronic Equipment*, Issue 1, May 2001.

Top500 Supercomputer Sites (2008). Top 10 Systems – 11/2008, http://www.top500.org

Trivedi K. S. (2001). *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*, ISBN 0-471-33341-7, John Wiley and Sons, New York