EVENT TRIGGERED CONTROL FOR CLOUD COMPUTING APPLICATIONS

A Project Report

submitted by

KSHAMA DWARAKANATH

in partial fulfilment of the requirements
for the award of the degree of

DUAL DEGREE (B.TECH. & M.TECH.)

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS.

MAY 2018

THESIS CERTIFICATE

This is to certify that the thesis titled **Event Triggered Control for Cloud Computing Applications**, submitted by **Kshama Dwarakanath**, to the Indian Institute of Technology Madras, for the award of the degree of **Dual Degree (B.Tech. & M.Tech)**, is a bona fide record of the research work done by her under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Ramkrishna Pasumarthy
Research Guide
Associate Professor
Dept. of Electrical Engineering
IIT-Madras, 600 036

Place: Chennai

Date: 11th May 2018

ACKNOWLEDGEMENTS

I would like to thank my guide, Dr. Ramkrishna Pasumarthy for giving me the opportunity to explore various research topics before choosing a problem of my interest. I am grateful to him for providing me with the resources to work on this problem. I would also like to thank Durgesh Singh with whom I have collaborated in this project. I am indebted to IIT Madras for providing me with a platform to study various courses and fields of research. Thanks are due to my parents and friends who have kept me motivated and focused through these five years at IIT Madras.

ABSTRACT

KEYWORDS: Event Triggered Control, Input to State Stability, $\mathcal{D}_R$ regions, Cloud Computing, Linear Matrix Inequality.

An event triggered control law is proposed for performance management of a web-server system hosted on a cloud while ensuring efficient resource utilization. The web server system instantiates or turn off Virtual Machines (VMs) in response to incoming web requests from clients. The aim of the control task is to allocate new VMs only as and when required, while guaranteeing that the CPU utilisation and the response time of the server system lie within certain predefined thresholds. The dynamics of the target system are modelled as a discrete linear time invariant system subject to exogenous inputs and actuation delay. Event triggering introduces measurement errors between triggering instants which are treated along with the exogenous inputs as 'disturbance input' to the system. The controller design problem is analysed within the framework of Input to State Stability (ISS) and $\mathcal{D}_R$ stability. A notion of 'Input to State $\mathcal{D}_R$ stability' is introduced, and an ISS Lyapunov function is designed to guarantee Input to State $\mathcal{D}_R$ stability of the system with respect to the disturbance input. An event triggering condition is derived for updating the control law whenever there is performance degradation or over provisioning of resources. The results are experimentally validated on a web-server system hosted in a private cloud environment.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

IITM	Indian Institute of Technology Madras
B.Tech.	Bachelor of Technology
M.Tech.	Master of Technology
ISS	Input to State Stable/Stability
VM	Virtual Machine
LMI	Linear Matrix Inequality
LTI	Linear and Time Invariant
KVM	Kernel-based Virtual Machine
TTC	Time Triggered Control
ETC	Event Triggered Control

CHAPTER 1

INTRODUCTION

One of the main objectives in any control application is to have optimal usage of resources, which is particularly relevant with the onset of systems with shared resources. Traditionally, control laws have been implemented in a periodic or a time-triggered manner making use of strong theoretical concepts that support it. Such a periodic update scheme results in excessive usage of computational and communication resources. Event triggered control (ETC) [12] is a technique where the control law is updated only when an event of interest has occurred. An event could be characterized by a deviation of the system from certain predefined performance metrics. Whilst ensuring performance, it is also necessary to make sure that resources are not over provisioned. Therefore, the event triggered control law must not only check for degradation in performance, but also trigger an event when resources are over provisioned. This event would cause a reduction in the control energy while maintaining quality of service. Since performance specifications can be mapped to specific pole regions for the closed loop system, efficient resource utilisation can be ensured by restricting the poles of the closed loop system to convex subregions of the complex plane, called $\mathcal{D}_R$ regions. There are two components to the design of such a controller for event triggering [10]: a feedback control law that determines the control input and a triggering mechanism that decides when the input is to be updated. The event triggering conditions are based on the measurement errors becoming large when compared with the state norm. However, in addition to measurement errors, most systems are also subject to exogenous inputs in terms of disturbances. Ensuring robustness of the system to disturbances and errors is dealt with by the notion of input-to-state (ISS) stability [18].

The authors in [15] use event triggering to reduce network traffic as well as energy expenditure of battery powered wireless sensor nodes in sensor/actuator networks. Event triggered control is used in [9] for distributed control of multi-agent systems where each agent is equipped with an embedded microprocessor with limited resources. This framework is preferable in such a scenario where each agent needs to have continuous communication with other agents. [14] deals with the design and implementation of a purely data-driven ETC technique for event detection and event handling in IoT applications. In [19], the problem of scheduling control tasks has been studied in the context of continuous time systems.

This project addresses the problem of event triggered control for a discrete time linear system subject to exogenous inputs. The specific system of interest is a web server system hosted on a private cloud. Server virtualisation in the cloud environment allows partitioning of the physical server into a number of virtual servers called Virtual Machines (VMs). Each VM hosts a web server which serves the incoming web requests. These VMs are instantiated (or turned off) based on the magnitude of the incoming workload and performance specifications. In such a set up, it is natural to use event triggered techniques since small changes in the incoming requests may not require the instantiation of a new VM. We derive an event triggering condition that monitors the states of the system and triggers a change in the control input only when certain performance specifications are violated, or when resources are over provisioned. The feedback control input is given by a proportional controller that ensures that closed loop poles are picked from a $\mathcal{D}_R$ region. A Lyapunov Function is designed to guarantee Input to State $\mathcal{D}_R$ stability for the web server system.

This project report is organised as follows. The concepts of Input to State Stability and $\mathcal{D}_R$ regions are introduced along with some mathematical preliminaries in Chapter 2. Chapter 3 presents the motivation behind this work along with the problem formulation. The design of an ISS-Lyapunov function to ensure Input to state $\mathcal{D}_R$ stability of the system with respect to the exogenous input as well as measurement errors introduced by event triggering is given in Chapter 4. Chapter 4 also covers the derivation of a theoretical event triggering condition based on the norm of the measurement error and exogenous input exceeding a certain threshold based on the states of the system. Chapter 5 presents the results from experimental validation of the event triggering scheme derived in Chapter 4 on a web server system hosted on a private cloud. It is observed that the control input is updated less frequently in event triggered control as compared to the case with time triggered control while meeting performance specifications. Moreover, over-provisioning of resources is prevented through the usage of a state feedback gain matrix that guarantees $\mathcal{D}_R$ stability of the closed loop system. The conclusions are presented in Chapter 6 along with some future work. These results have been submitted to the 57th IEEE Conference on Decision and Control 2018 [4].

CHAPTER 2

PRELIMINARIES

Definition 1. *A continuous function $\alpha : [0,a) \to [0,\infty)$ belongs to class $\mathcal{K}$ if it is strictly increasing and $\alpha(0) = 0$. It belongs to class $\mathcal{K}_\infty$ if $a = \infty$ and $\alpha(r) \to \infty$ as $r \to \infty$.*

Definition 2. *A continuous function $\beta : [0,a) \times [0,\infty) \to [0,\infty)$ belongs to class $\mathcal{KL}$ if*

1. *for each s, $\beta(r,s)$ belongs to class $\mathcal{K}$ with respect to r and,*
2. *for each r, $\beta(r,s)$ decreases with respect to s and $\beta(r,s) \to 0$ as $s \to \infty$.*

Definition 3. *[8] The system*

$$x(k+1) = Ax(k) + Bu(k) \tag{2.1}$$

is said to be Input to State Stable (ISS) with respect to input u if there exist functions $\beta \in \mathcal{KL}$ and $\gamma \in \mathcal{K}$ such that for all $u \in \mathbb{R}^p$ and $x(0) \in \mathbb{R}^n$

$$\|x(k)\| \leq \beta(\|x(0)\|, k) + \gamma(\|u\|_k),\ \forall k \in \mathbb{N} \tag{2.2}$$

with $\|u\|_k = \sup\{\|u(k)\| : k \in \{0, 1, \cdots, k-1\}\}$.

Definition 4. *A continuous function $V : \mathbb{R}^n \to \mathbb{R}^+$ is said to be an ISS-Lyapunov Function for the system (2.1) if there exist functions $\alpha_1, \alpha_2 \in \mathcal{K}_\infty$ such that*

$$\alpha_1(\|x\|) \leq V(x) \leq \alpha_2(\|x\|),\ \forall x \in \mathbb{R}^n \tag{2.3}$$

and functions $\alpha \in \mathcal{K}_\infty$ and $\chi \in \mathcal{K}$ such that

$$V(x(k+1)) - V(x(k)) \leq -\alpha(\|x(k)\|) + \chi(\|u(k)\|) \tag{2.4}$$

The existence of such an ISS-Lyapunov function guarantees that the system (2.1) is ISS with respect to u.

Definition 5. *[6] A Linear Matrix Inequality (LMI) has the form*

$$F(x) := F_0 + \sum_{i=1}^{m} x_t F_i > 0 \tag{2.5}$$

where $x \in \mathbb{R}^m$ is the variable and the symmetric matrices $F_i = F_i^T \in \mathbb{R}^{n\times n}$, $i = 0, 1, \vdots, m$ are known. The inequality $F(x) > 0$ means that $F(x)$ is positive definite, hence, $v^T F(x) v > 0,$ for all *$v \neq 0$. The linear matrix inequalities could have matrices are the variables, as in this work.*

Lemma 1. *[5] Schur Complement: Consider a symmetric matrix $M \in \mathbb{R}^{(p+q)\times(p+q)}$ of the form*

$$M = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$$

where $A \in \mathbb{R}^{p\times p}$, $B \in \mathbb{R}^{p\times q}$, $C \in \mathbb{R}^{q\times q}$ and C is invertible. Then,

$$M > 0 \iff C > 0,\ A - BC^{-1}B^T > 0$$

Definition 6. *[16] A $\mathcal{D}_R$ region in the complex plane is defined by the symmetric matrix*

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{12}' & R_{22} \end{bmatrix} \tag{2.6}$$

as

$$\mathcal{D}_R = \{z \in \mathbb{C} : R_{11} + R_{12}z + R_{12}'z^\star + R_{22}zz^\star < 0\} \tag{2.7}$$

where $R \in \mathbb{R}^{2d\times 2d}$, $R_{11} = R_{11}^T \in \mathbb{R}^{d\times d}$, $R_{12} \in \mathbb{R}^{d\times d}$ and $R_{22} = R_{22}^T \in \mathbb{R}^{d\times d}$.

Definition 7. *The discrete time system $x(k+1) = Ax(k)$ is said to be $\mathcal{D}_R$ stable if all the eigen values of A lie in the $\mathcal{D}_R$ region given by (2.7).*

Definition 8. *The system*

$$x(k+1) = Ax(k) + Bu(k)$$

is said to be Input to State $\mathcal{D}_R$ stable with respect to u if:

i. The system is ISS with respect to u.

ii. The eigen values of A lie in a $\mathcal{D}_R$ region.

We are interested in a $\mathcal{D}_R$ region that is a disc in the complex plane with center $c_D \in \mathbb{R}$ and radius $r_D > 0$ defined by the matrix

$$R = \begin{bmatrix} -r_D & -c_D & 0 & 1 \\ -c_D & -r_D & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{2.8}$$

The $\mathcal{D}_R$ region is plotted for $r_D = 0.5$ and $c_D = 0.25$ in Figure 2.1. The shaded area is the $\mathcal{D}_R$ region with the outer unit circle denoting the region of stability for discrete time systems.

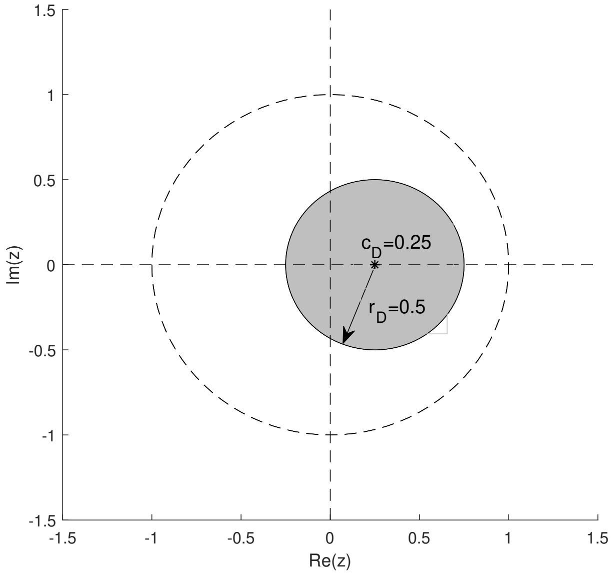

Figure 2.1: Plot of $\mathcal{D}_R$ region with $r_D = 0.5$ and $c_D = 0.25$

Lemma 2. *[7] The system $x(k+1) = Ax(k)+Bu(k)$ is $\mathcal{D}_R$ stable in the region defined by (2.8) whist being asymptotically stable under the state feedback law $u(k) = Kx(k)$ if there exist matrices $U = U^T > 0$ and V such that*

$$\begin{bmatrix} r_D U & c_D U - AU - BV & 0 & 0 \\ \star & r_D U & 0 & 0 \\ \star & \star & U & -AU - BV \\ \star & \star & \star & U \end{bmatrix} > 0$$

The feedback matrix K is then given as $K = VU^{-1}$.

CHAPTER 3

MOTIVATION AND PROBLEM FORMULATION

The operation of the web server system hosted on the cloud in the presence of a time triggered controller is analysed. In this case, the number of VMs in operation is updated every sampling instant in response to the incoming workload. The response time and CPU utilization of the system are plotted along with the control input for slow varying workload in Figure 3.1 and for fast varying workload in Figure 3.2.

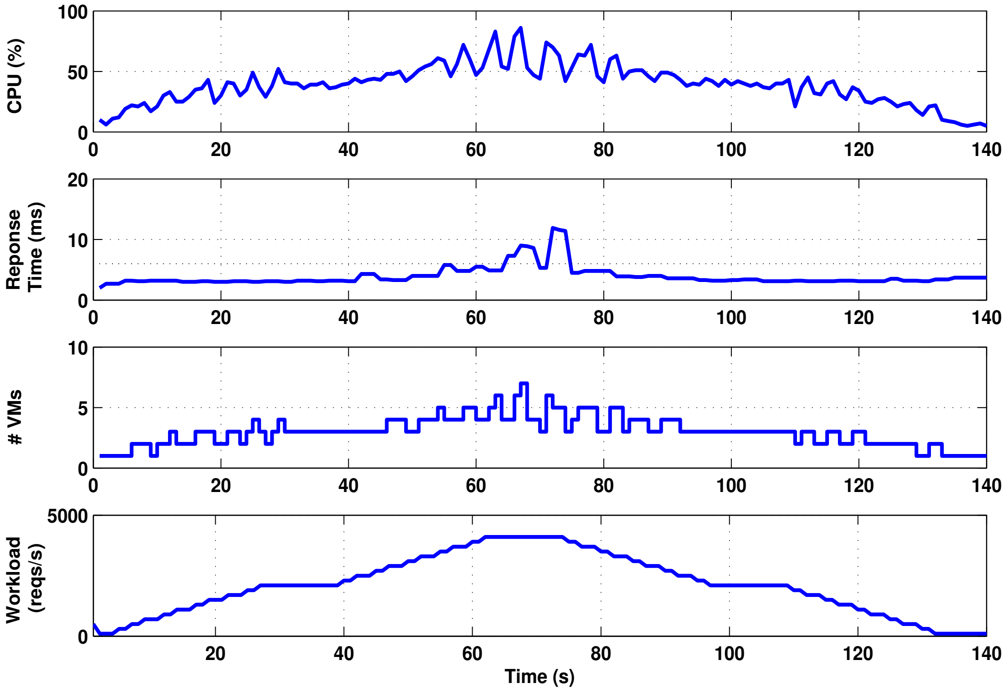

Figure 3.1: Time Triggered Control with slow varying workloads

From Figure 3.1, it is observed that small changes in workload do not require corresponding change in the control law over subsequent sampling instants. Hence, the opening up the possibility to reduce the frequency of communication between actuators and the system by keeping the control law constant until certain performance thresholds are violated. In the case of fast varying workload as shown in Fig. 3.2, we observe that time triggered control action is more rapid. These cases motivate us to adapt to an event triggered control approach that would decrease the frequency of control updates resulting in lesser control effort and lower bandwidth consumption.

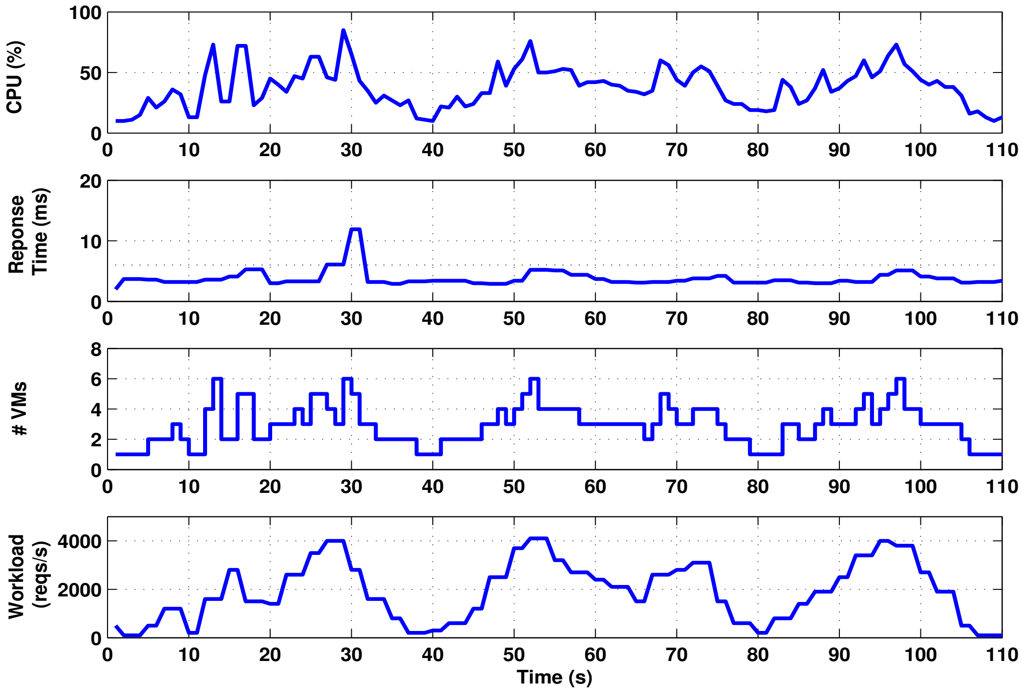

Figure 3.2: Time Triggered Control with fast varying workloads

3.1 Problem Formulation

The web-server system on the cloud is modelled as a discrete linear and time invariant (LTI) system with states as the mean physical core utilization denoted by $CPU(k)$ (in %), and the mean web-request response time denoted by $RES(k)$ (in ms). These can be represented by the state vector $x(k) = \begin{bmatrix} CPU(k) & RES(k) \end{bmatrix}^T$. The system is subject to the exogenous input of incoming web-request rate $u_2(k)$ measured in requests per second. The control input is the number of web-servers or virtual machines handling the incoming web-requests and is denoted by $u_1(k)$. The dynamics of the system are given as

$$x(k+1) = Ax(k) + Bu_1(k) + Wu_2(k) \tag{3.1}$$

where $x(k) \in \mathbb{R}^2$, $u_1(k) \in \mathbb{R}^1$, $u_2(k) \in \mathbb{R}^1$, $A \in \mathbb{R}^{2\times2}$, $B \in \mathbb{R}^{2\times1}$ and $W \in \mathbb{R}^{2\times1}$. These matrices are determined using system identification techniques as in [17] around an operating region defined by the incoming workload. We assume that the incoming workload $u_2(k)$ is bounded:

$$0 < r \leq u_2(k) \leq R, \, \forall k \in \{0, 1, 2, ...\} \tag{3.2}$$

The aim is to design an event triggered controller that comprises of a state feedback control law, and a triggering condition for the update of the control input; whilst ensuring Input to State $\mathcal{D}_R$ stability of the system. The states of the system are continuously monitored and tested against the triggering condition. Let the time instants of triggering be k_0, k_1, ... and the actuation instants be $k_0+\Delta$, $k_1+\Delta$, ... as in Fig 3.3. Here, $\Delta \in \mathbb{Z}^+$ is the time taken for state measurement, computation and update of the control law.

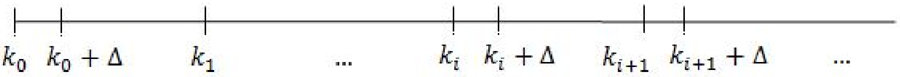

Figure 3.3: Time instants of triggering and control update

The control input is held constant between two updates, i.e,

$$u_1(k) = Kx(k_i),\ \forall k \in [k_i+\Delta, k_{i+1})$$

Define *measurement error* in the interval between two updates $[k_i+\Delta, k_{i+1})$ as

$$e(k) = x(k_i) - x(k)$$

The closed loop system is given by

$$\begin{aligned} x(k+1) &= Ax(k) + BKx(k_i) + Wu_2(k) \\ &= (A+BK)x(k) + BKe(k) + Wu_2(k) \end{aligned}$$

Denote by $v(k) = \begin{bmatrix} e^T(k) & u_2^T(k) \end{bmatrix}^T$ the stack of measurement errors as well as exogenous inputs, which are treated equivalently as disturbance to the system. Let $\tilde{A} := A+BK$ and $E := \begin{bmatrix} BK & W \end{bmatrix}$ so that the evolution of the closed loop system is given by

$$x(k+1) = \tilde{A}x(k) + Ev(k) \tag{3.3}$$

The problem statement is as follows:

Use event triggered control to achieve reduction in resource consumption of a system subject to exogenous inputs and actuation delay. Performance specifications on the states of the system are to be met while ensuring that there is no over provisioning of resources.

In order to achieve this, the following steps are followed:

1. Design an ISS Lyapunov function to guarantee input to state $\mathcal{D}_R$ stability of the system in (3.3) with respect to the disturbance input v.

2. Derive an event triggering condition to update the control input using the Lyapunov function obtained above.

3. Test the event triggering condition on the web server system hosted on the private cloud.

CHAPTER 4

INPUT TO STATE $\mathcal{D}_R$ STABILITY AND CONTROLLER DESIGN

4.1 Design of Lyapunov Function

The aim is to find a quadratic Lyapunov function of the form

$$V(x) = x^T P x \tag{4.1}$$

with $P = P^T > 0$ that renders the closed loop system (3.3) to be Input to State $\mathcal{D}_R$ stable with respect to v.

Theorem 1. *The system*

$$\begin{aligned} x(k+1) &= (A+BK)x(k) + Ev(k) \\ &= \tilde{A}x(k) + Ev(k) \end{aligned}$$

is Input to State $\mathcal{D}_R$ stable with respect to v if there exists a solution $P = P^T > 0$ to the LMI

$$\begin{bmatrix} \gamma I & 0 & E^T P & 0 \\ 0 & P & -\tilde{A}^T P & Q \\ PE & -P\tilde{A} & P & 0 \\ 0 & Q & 0 & Q \end{bmatrix} > 0 \tag{4.2}$$

given a feedback gain matrix K that renders $\tilde{A}$ to be $\mathcal{D}_R$ stable.

Proof. For the given system to be Input to State $\mathcal{D}_R$ Stable using Definition 8, it is required that the system be Input to State Stable with respect to v given a feedback gain matrix K that ensures that $\tilde{A}$ is $\mathcal{D}_R$ stable. We wish to find a $P = P^T > 0$ such that the quadratic Lyapunov function in (4.1) is an ISS-Lyapunov function for the system. We have

$$\lambda_{\min}(P)\|x\|^2 \leq V(x) \leq \lambda_{\max}(P)\|x\|^2 \tag{4.3}$$

where $\alpha_1(x) := \lambda_{\min}(P)\|x\|^2$ and $\alpha_2(x) := \lambda_{\max}(P)\|x\|^2$ are $\mathcal{K}_\infty$ functions as in (2.3).

$$\begin{aligned} V(x(k+1)) - V(x(k)) &= x^T(k+1)Px(k+1) - x^T(k)Px(k) \\ &= x^T\left(\tilde{A}^T P\tilde{A} - P\right)x + 2x^T\left(\tilde{A}^T PE\right)v + v^T\left(E^T PE\right)v \end{aligned}$$

We wish to find a P such that

$$V(x(k+1)) - V(x(k)) \leq -x^T Qx + \gamma\|v\|^2 \tag{4.4}$$

for some $Q = Q^T > 0$ and $\gamma > 0$ as in (2.4). Define

$$\begin{aligned} f(x,v) &:= x^T\left(\tilde{A}^T P\tilde{A} - P + Q\right)x + 2x^T\left(\tilde{A}^T PE\right)v + v^T\left(E^T PE - \gamma I\right)v \\ &= -x^T Sx + 2x^T\left(\tilde{A}^T PE\right)v + v^T\left(E^T PE - \gamma I\right)v \end{aligned}$$

where $\tilde{A}^T P\tilde{A} - P + Q := -S$ with $S = S^T > 0$. For the system to be ISS with respect to v, it is required that

$$f(x,v) \leq 0,\ \forall x \in \mathbb{R}^2,\ \forall v \in \mathbb{R}^3$$

Since $f(x,v)$ is a concave function of x, it can be ensured that $f(x,v) \leq 0,\ \forall x \in \mathbb{R}^2,\ \forall v \in \mathbb{R}^3$ by enforcing $f(x^\star(v),v) \leq 0,\ \forall v \in \mathbb{R}^3$ where $x^\star(v)$ is the point of maximum.

$$\frac{\partial f(x,v)}{\partial x} = -2Sx + 2\tilde{A}^T PEv = 0 \Rightarrow x^*(v) = S^{-1}\tilde{A}^T PEv$$

Note that $\frac{\partial^2 f(x,v)}{\partial x^2} = -2S < 0$ so that $x^*(v)$ is indeed a point of maximum.

$$f(x^*(v),v) = v^T\left(E^T PE + E^T P\tilde{A}S^{-1}\tilde{A}^T PE - \gamma I\right)v$$

Hence, ISS requires that

$$\gamma I - E^T PE - E^T P\tilde{A}S^{-1}\tilde{A}^T PE > 0$$

Using Lemma 1 and expanding for S, we require that

$$\begin{bmatrix} \gamma I - E^T PE & E^T P\tilde{A} \\ \tilde{A}^T PE & P - Q - \tilde{A}^T P\tilde{A} \end{bmatrix} > 0 \tag{4.5}$$

which is the same as

$$\begin{bmatrix} \gamma I & 0 \\ 0 & P \end{bmatrix} - \begin{bmatrix} E^T P & 0 \\ -\tilde{A}^T P & Q \end{bmatrix} \begin{bmatrix} P^{-1} & 0 \\ 0 & Q^{-1} \end{bmatrix} \begin{bmatrix} PE & -P\tilde{A} \\ 0 & Q \end{bmatrix} > 0$$

The above inequality gives rise to (4.2) using Lemma 1. □

4.2 Event Triggering Condition

Theorem 1 is used to design an ISS Lyapunov function that ensures Input to State $\mathcal{D}_R$ stability of system (3.3) with respect to disturbance input v. An event triggering condition is now derived using this Lyapunov function.

Theorem 2. *Given the state $x(k)$ and the disturbance input $v(k)$ to the system (3.3) at any instant k, we have the following triggering conditions based on the delay Δ:*

1. *When $\Delta = 0$:*
$$\|v(k)\| \leq \mu \|x(k)\| \tag{4.6}$$
where
$$\mu = \left(\frac{\sigma \lambda_{min} \left(P - \tilde{A}^T P \tilde{A} \right)}{\lambda_{max} \left(E^T P E \right) + 2\alpha_v \|E\| \|P\| \|\tilde{A}\|} \right)^{0.5}$$

2. *When $\Delta \geq 1$:*
$$\|v(k)\| \leq \mu' \|x(k)\| \tag{4.7}$$
where
$$0 < \frac{r}{R \left[\alpha_v a^{\Delta} + \left(\frac{1-a^{\Delta}}{1-a} \right) \left(\alpha_v \|BK\| + \|W\| \right) \right]} \leq \mu' < \mu, \tag{4.8}$$
where $a := \|A\|$.

In both cases, $0 < \sigma < 1$ and it is assumed that $\|\tilde{A}\| < 1$[1]*. The above condition ensures Input to State $\mathcal{D}_R$ stability of the system (3.3) with respect to v.*

Proof. The first case considered is when $\Delta = 0$. Triggering occurs whenever there is a violation of Input to State $\mathcal{D}_R$ stability condition, which can be checked as follows.

$$V(x(k+1)) - V(x(k)) = x^T \left(\tilde{A}^T P \tilde{A} - P \right) x + 2x^T \left(\tilde{A}^T P E \right) v + v^T \left(E^T P E \right) v$$

[1] This condition on the closed loop system matrix is satisfied for our web server system.

Since

$$
\begin{aligned}
x^T\left(\tilde{A}^T P\tilde{A}-P\right)x &\leq -\lambda_{min}\left(P-\tilde{A}^T P\tilde{A}\right)\|x\|^2\\
v^T\left(E^T PE\right)v &\leq \lambda_{max}\left(E^T PE\right)\|v\|^2\\
2v^T\left(E^T P\tilde{A}\right)x &\leq 2\|E\|\|P\|\|\tilde{A}\|\|v\|\|x\|
\end{aligned}
$$

we have

$$
\begin{aligned}
V(x(k+1))-V(x(k)) &\leq -\lambda_{min}\left(P-\tilde{A}^T P\tilde{A}\right)\|x\|^2+\lambda_{max}\left(E^T PE\right)\|v\|^2\\
&\quad +2\|E\|\|P\|\|\tilde{A}\|\|v\|\|x\| \qquad (4.9)
\end{aligned}
$$

The closed loop system is given by $x(k+1)=\tilde{A}x(k)+Ev(k)$. Hence,

$$
x(k)=\tilde{A}^k x(0)+\sum_{j=0}^{k-1}\tilde{A}^{k-1-j}Ev(j)
$$

Define $\|v\|_l := \min\{\|v(j)\|, j=0,1,...k-1\}$ and $\|v\|_h := \max\{\|v(j)\|, j=0,1,...k-1\}$. For a bounded workload of the form as in (3.2), we have $\|v\|_l = r$ and a finite $\|v\|_h$. Therefore,

$$
\begin{aligned}
\|x(k)\| &\leq \|\tilde{A}\|^k\|x(0)\|+\sum_{j=0}^{k-1}\|\tilde{A}\|^{k-1-j}\|E\|\|v(j)\|\\
&\leq \|\tilde{A}\|^k\|x(0)\|+\left(\frac{1-\|\tilde{A}\|^k}{1-\|\tilde{A}\|}\right)\|E\|\|v\|_h\\
&= \left[\frac{\|\tilde{A}\|^k\|x(0)\|}{\|v(k)\|}+\left(\frac{1-\|\tilde{A}\|^k}{1-\|\tilde{A}\|}\right)\frac{\|E\|\|v\|_h}{\|v(k)\|}\right]\|v(k)\|\\
&\leq \left[\frac{\|\tilde{A}\|^k\|x(0)\|}{\|v\|_l}+\left(\frac{1-\|\tilde{A}\|^k}{1-\|\tilde{A}\|}\right)\frac{\|E\|\|v\|_h}{\|v\|_l}\right]\|v(k)\|
\end{aligned}
$$

Under the assumption that $\|\tilde{A}\|<1 \Rightarrow \|\tilde{A}\|^k<1,\ \forall\, k>1$,

$$
\|x(k)\| \leq \alpha_v\|v(k)\| \qquad (4.10)
$$

where

$$
\alpha_v := \frac{\|x(0)\|}{r}+\frac{\|E\|\|v\|_h}{r\left(1-\|\tilde{A}\|\right)}
$$

Using (4.10) in (4.9),

$$\begin{aligned} V(x(k+1)) - V(x(k)) &\leq -\lambda_{min}\left(P - \tilde{A}^T P \tilde{A}\right)\|x\|^2 \\ &\quad + \left(\lambda_{max}\left(E^T P E\right) + 2\alpha_v \|E\|\|P\|\|\tilde{A}\|\right)\|v\|^2 \\ &\leq -\alpha(\|x\|) + \chi(\|v\|) \end{aligned}$$

which is the same as equation (2.4) with $\alpha(\|x\|) = \lambda_{min}\left(P - \tilde{A}^T P \tilde{A}\right)\|x\|^2$ and $\chi(|v\|) = \left(\lambda_{max}\left(E^T P E\right) + 2\alpha_v \|E\|\|P\|\|\tilde{A}\|\right)\|v\|^2$. Using a triggering condition of the form $\chi(\|v\|) \leq \sigma\alpha(\|x\|)$ where $0 < \sigma < 1$, we get (4.6).

In the case when $\Delta \geq 1$, the triggering must happen before condition (4.6) is violated in order to compensate for the delay in actuation. Hence, the triggering condition is of the form $\|v(k)\| \leq \mu'\|x(k)\|$ with $\mu' < \mu$, where μ represents the triggering threshold without delay (when $\Delta = 0$). Let the controller be triggered at $k = k_i$ due to $\|v(k)\| \geq \mu'\|x(k)\|$. In order to ensure that there is no triggering during the period of actuation $k \in (k_i, k_i + \Delta)$, we require that $\|v(k_i+\Delta)\| \leq \mu'\|x(k_i+\Delta)\|$. This requires that

$$\begin{aligned} \mu' &\geq \frac{\|v(k_i+\Delta)\|}{\|x(k_i+\Delta)\|} \\ &\geq \frac{\left\|\begin{bmatrix} e(k_i+\Delta) \\ u_2(k_i+\Delta) \end{bmatrix}\right\|}{\|x(k_i+\Delta)\|} \\ &\geq \frac{\|u_2(k_i+\Delta)\|}{\|x(k_i+\Delta)\|} \\ &\geq \frac{r}{\|x(k_i+\Delta)\|} \text{ (using (3.2))} \end{aligned} \tag{4.11}$$

The evolution of the system for $k \in (k_i, k_i + \Delta)$ is given by

$$x(k+1) = Ax(k) + c_1 + Wu_2(k)$$

where $c_1 := Bu_1(k_i) = BKx(k_{i-1})$ is a constant over the period $(k_i, k_i + \Delta)$. The above difference equation is solved for $x(k_i + \Delta)$ to obtain

$$x(k_i+\Delta) = A^\Delta x(k_i) + \left(\sum_{j=0}^{\Delta-1} A^j\right) c_1 + \left(\sum_{j=0}^{\Delta-1} A^j W u_2(k_i+\Delta-1-j)\right)$$

Using (4.10),

$$\begin{aligned}
\|x(k_i)\| &\leq \alpha_v \|v(k_i)\| \\
&\leq \alpha_v \| \begin{bmatrix} 0 \\ u_2(k_i) \end{bmatrix} \| \\
&\leq \alpha_v R \\
\|c_1\| &= \|BKx(k_{i-1})\| \\
&\leq \alpha_v R \|BK\|
\end{aligned}$$

Define $a := \|A\|$ to get

$$\|x(k_i + \Delta)\| \leq R \left[\alpha_v a^{\Delta} + \left(\frac{1 - a^{\Delta}}{1 - a} \right) (\alpha_v \|BK\| + \|W\|) \right] \tag{4.12}$$

Using (4.12) in (4.11) gives (4.8). □

Remark 1. *When the actuation delay $\Delta = 0$, the event triggering condition (4.6) is shown to be non-trivial in Proposition 2 of [11]. The Proposition guarantees the absence of a Zeno like phenomenon for discrete time systems, which means that any two triggering time instants are at least two time steps apart. On the other hand, the presence of non-zero delay trivially prevents the occurrence of such a phenomenon.*

CHAPTER 5

EXPERIMENTAL VALIDATION

5.1 Experimental Setup

The experimental setup consists of three physical machines - the server machine, the client machine and the controller machine. The server machine is used to host the web services. It has 32 GB of physical memory and 2 sockets, each containing 4 cores clocked at 3.2 GHz. Kernel-based Virtual Machine (KVM) [2] virtualization software (version qemu-kvm-1.0) is used on this machine to create and manage VMs. Each VM instance is allocated one physical core, 2 GB of memory and is deployed on the Apache web server to host a static web page. The server machine executes the HAProxy [1] load balancer which uses a round robin policy to distribute incoming workload equally among the active VMs in the system. The mean CPU utilization of the VMs is sampled every 5 seconds at the server machine to facilitate control.

The client machine has 16 GB of physical memory and 4 cores clocked at 3.2 GHz. It executes the Httperf [13] workload generator to submit a synthetic workload to the VMs hosted on the server machine. The instrumentation provided by Httperf is used to record the mean response time of the VMs every 5 seconds.

MATLAB [3] is used to implement the controller on the third machine which is used as a controller machine. The mean response time received from the client machine and the mean CPU utilization received from the server machine are provided as inputs to the controller. The controller outputs the control input corresponding to the number of active VMs required to serve the incoming web requests. This is sent to the server machine. These three machines are connected to each other using a dedicated 1 Gbps Ethernet switch. A schematic of the experimental set-up of the web server system on the cloud is shown in Figure 5.1.

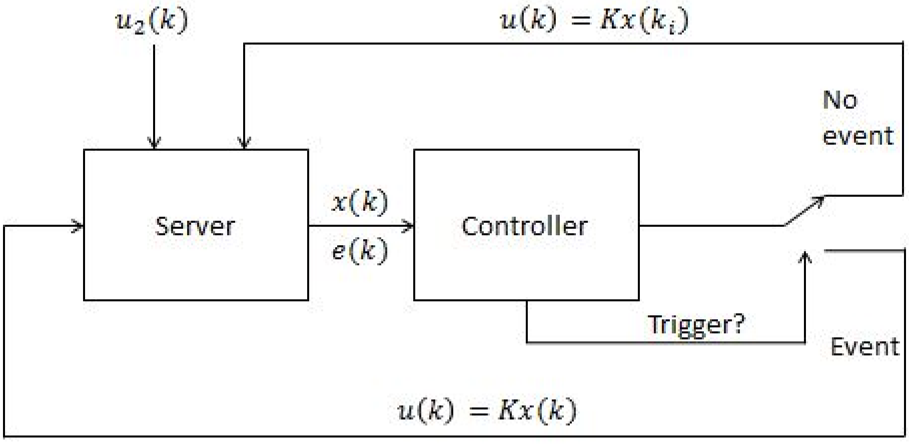

Figure 5.1: Schematic of set-up of web server system on cloud

5.2 System Identification

The web server system hosted on the cloud is modelled as a discrete linear time invariant system with two states $x(k) = \begin{bmatrix} CPU(k) & RES(k) \end{bmatrix}^T$ where $CPU(k)$ is the mean CPU utilisation in % and $RES(k)$ is the mean response time in ms. The following model is experimentally determined using the system identification techniques explained in [17]:

$$\begin{aligned} \begin{bmatrix} CPU(k+1) \\ RES(k+1) \end{bmatrix} &= \begin{bmatrix} 0.9300 & 0.0941 \\ 0.0141 & 0.8226 \end{bmatrix} \begin{bmatrix} CPU(k) \\ RES(k) \end{bmatrix} + \begin{bmatrix} -0.3873 \\ -0.1332 \end{bmatrix} VM(k) \\ &+ \begin{bmatrix} 0.0018 \\ 0.0003 \end{bmatrix} u_2(k) \end{aligned} \tag{5.1}$$

Here, $VM(k)$ denotes the control input to the system which is the number of active Virtual Machines, and $u_2(k)$ is the incoming workload in web requests per second. The dynamics in (5.1) are derived over an operating range defined by

$$1req/s \leq u_2(k) \leq 40000req/s$$

The system matrices in (5.1) are used in Lemma 2 to compute the state feedback gain matrix K such that the closed loop poles lie in a $\mathcal{D}_R$ region centred at $c_D = 0.85$ with radius $r_D = 0.1$. Further, an ISS Lyapunov function is designed by solving the Linear Matrix Inequality in Theorem 1 to obtain the Lyapunov matrix P. Finally, the event triggering bound μ is derived for $\Delta = 0$ as in (4.6), and the triggering bound μ' is derived for $\Delta = 2$ as in (4.7). The results of the experiments are given in the next section.

5.3 Results

The performance of event triggered control is compared with that of time triggered control of the web-server system for the cases when $\Delta = 0$ and $\Delta = 2$. In both these techniques, the control input is computed using the state feedback gain matrix K derived in the previous section. The update of the control input happens at every sampling instant for time triggered control. In the case of event triggered control, the states of the system $x(k)$ and the disturbance input $v(k)$ are continuously measured and tested against the event triggering condition. The triggering condition is said to be active and holds a value of '1' if an event occurs, requiring an update of the control law. It is said to be inactive otherwise and holds a value of '0'. The feedback control law used renders the system Input to State $\mathcal{D}_R$ stable with respect to the disturbance input v.

In order to test the efficacy of the theoretically derived event triggering condition in (4.6) or (4.7) as the case maybe, its performance is compared against that of event triggering with practical bounds on the system states. The chosen $\mathcal{D}_R$ region corresponds to the following bounds on the states of the system:

$$t_{low} \leq RES \leq t_{high} \tag{5.2}$$

$$CPU \leq CPU_{high} \tag{5.3}$$

where $t_{low} = 1.5ms$, $t_{high} = 2ms$ and $CPU_{high} = 35\%$. The upper bounds on the response time and CPU utilisation correspond to performance specifications of the controller while the lower bound on the response time enforces efficient resource utilisation. For example, a sudden increase in workload would require an update of the control law in order to meet the performance specifications. Thus, resulting in an increase in the number of active VMs. Now, if the system experiences a decrease in the workload, the event triggering condition would not update the control law in the absence of a lower

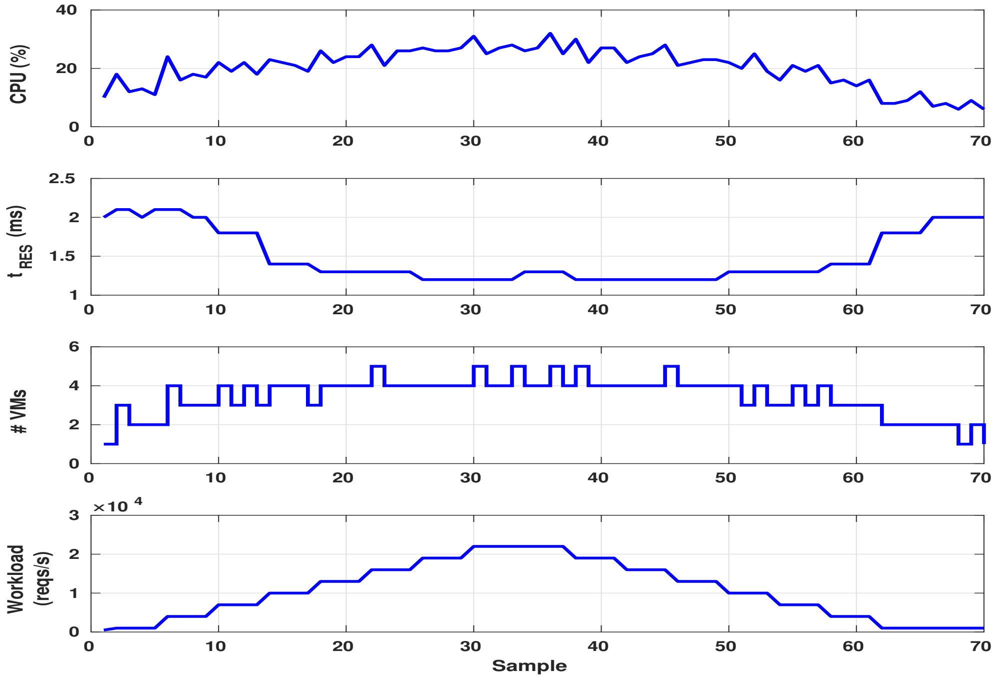

Figure 5.2: Time Triggered Control with zero delay

bound on response time. This results in the operation of more VMs than are necessary. Hence, the choice of the state feedback gain matrix K to ensure $\mathcal{D}_R$ stability deals with the issue of over provisioning of resources.

Figure 5.2 shows the evolution of CPU utilisation and response time of the server system when time triggered control is used. The incoming workload increases linearly with time and then, decreases. In this case, the actuation delay is assumed to be zero resulting in an immediate update of the control input. The figure also shows the control input (number of VMs) and the incoming workload. The control input is seen to increase with increase in workload, and decrease as the workload decreases. It is also observed that the control input remains the same at many instants, while the computation of the control law happens at each sampling instant. Hence, the network bandwidth consumed is high due to frequent updates of control law while the value of the control input may remain the same over sub-sequent instants. It is also observed that the CPU utilisation increases as the workload increases, and decreases accordingly whilst meeting the performance specifications. A key observation is that the response time decreases in some cases when the workload increases since there are more number of VMs serving the web-requests.

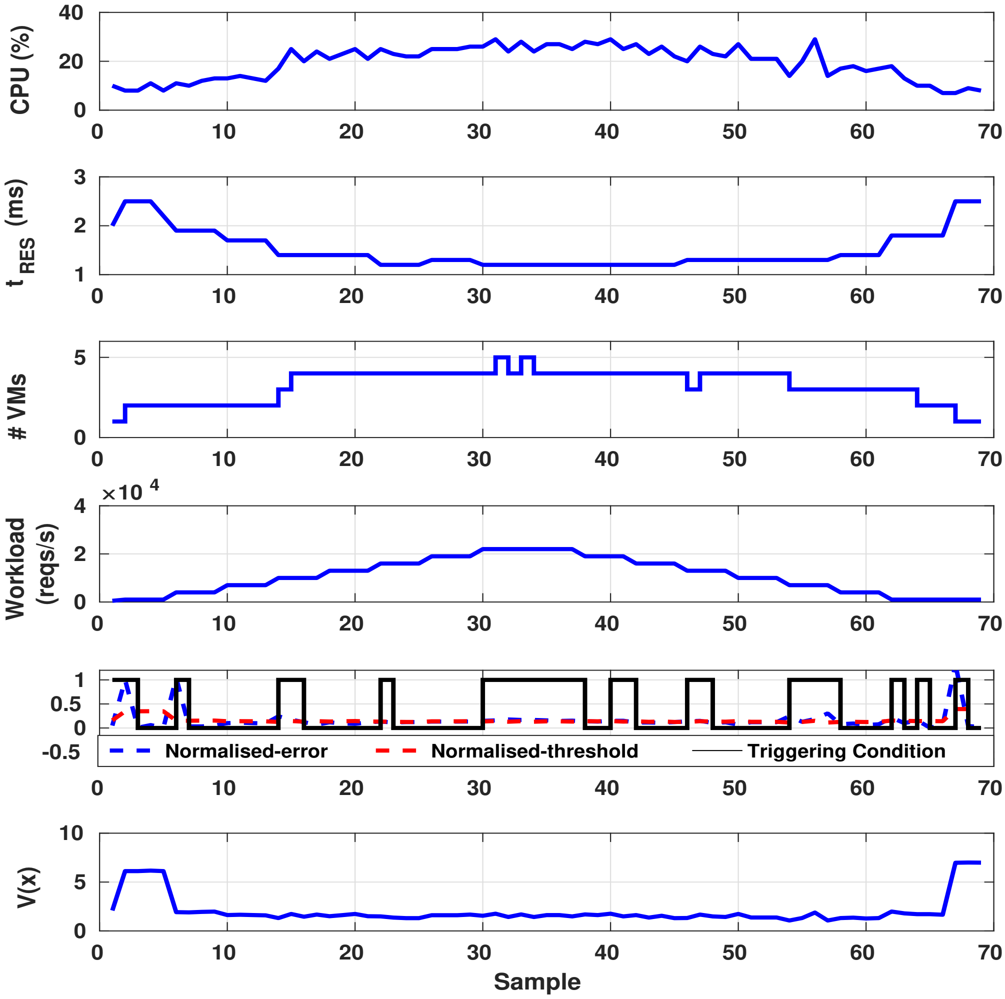

Figure 5.3: Event Triggered Control with theoretical bounds with zero delay

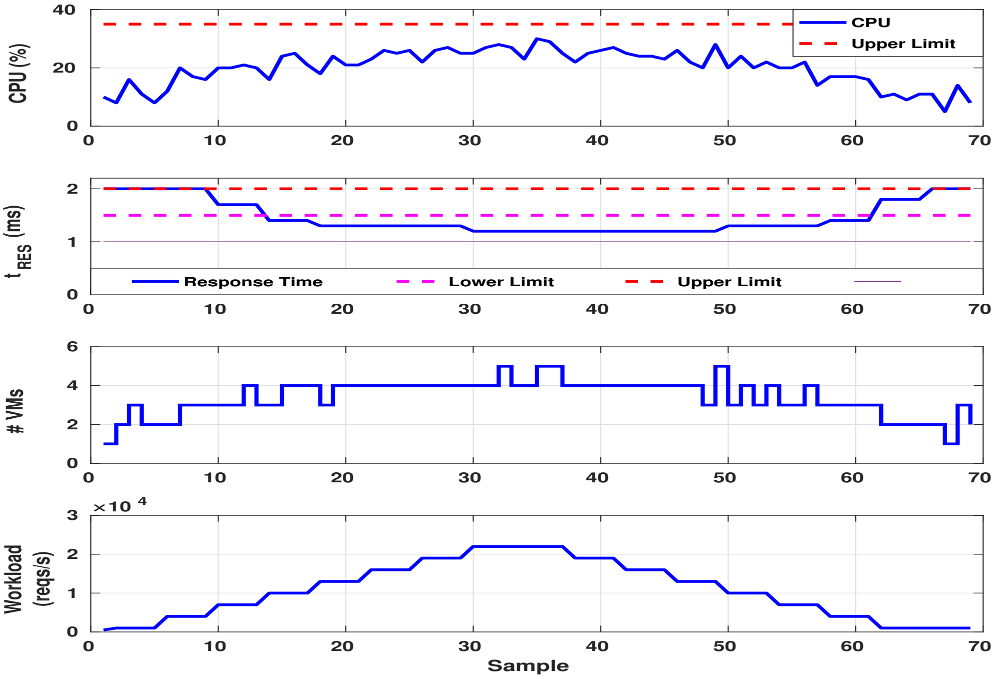

Figure 5.4: Event Triggered Control with practical bounds with zero delay

Figure 5.3 shows the evolution of the system states along with the control input, the workload, the triggering condition and the ISS-Lyapunov function with event triggered control using the triggering condition in (4.6) with zero delay. The workload increases linearly with time and then, decreases. The control law is updated whenever an event occurs, denoted by the triggering condition taking the value ‘1‘. The control input is held constant when the triggering condition is ‘0‘. It is observed that the updates are required less often than in the case of time triggered control, while satisfying performance specifications. A decrease in the control input is observed as the workload decreases due to the usage of $\mathcal{D}_R$ regions. Thereby, preventing over-provisioning of resources. This approach results in lower bandwidth consumption as well as efficient resource utilisation.

Figure 5.4 shows the evolution of the system states along with the control input and the workload, with event triggered control using practical bounds on the system states as in (5.2) and (5.3) with zero delay. The workload increases linearly with time and then, decreases. The control law is updated whenever the response time goes outside the desired range or when the CPU utilisation goes above its upper threshold. It is observed that while the specification on CPU utilisation is always met, the response time tends to go lower than required due to larger number of VMs than necessary. Whenever such an

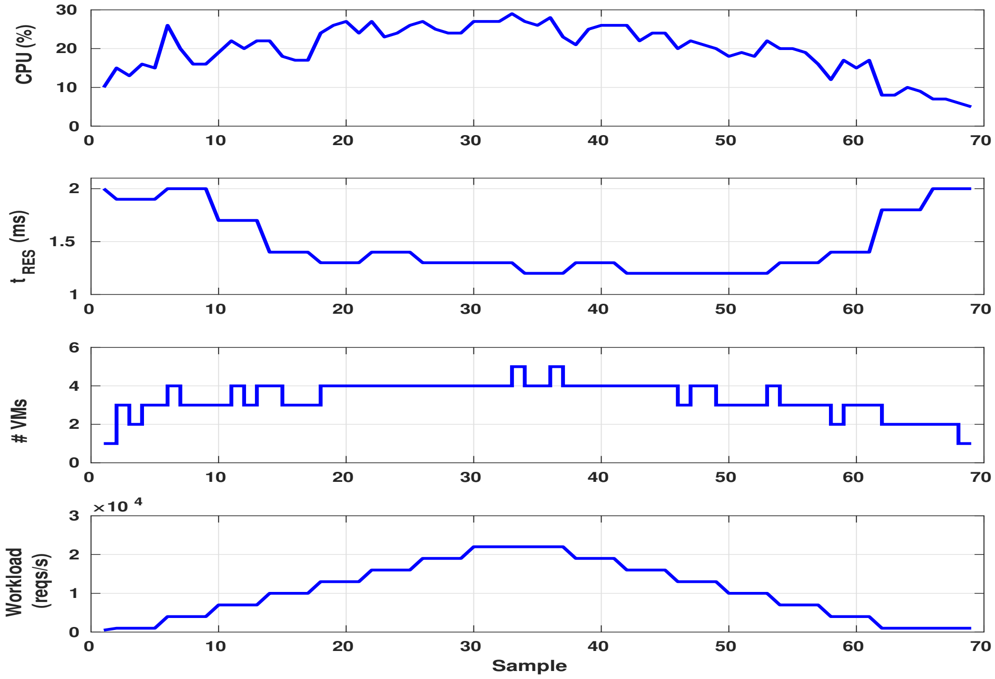

Figure 5.5: Time Triggered Control with $\Delta = 2$

event occurs, it is seen that the control input tends to decrease in the future instants to bring the response time back within the desired range. The control input is held constant when there is no event. Once again, this approach results in less frequent updates of the control input as compared to time triggered control. At the same time, the performance of event triggering with the triggering condition in (4.6) is better than that in this case.

The next set of experiments compare the performance of time triggered and event triggered control of the web server system in the presence of an actuation delay of two sampling instant, i.e, $\Delta = 2$. In the case of time triggered control, though the control input is computed at every time instant, the update happens only two instants later. With event triggered control, the update happens two instants after the triggering instant. after triggering occurs. The incoming workload is the same as that in the previous experiments. It increases linearly with time and then, decreases.

Figure 5.5 shows the evolution of CPU utilisation and response time of the server system with time triggered control for $\Delta = 2$. The figure also shows the control input (number of VMs) and the incoming workload. The control input is seen to increase with increase in workload, and decrease as the workload decreases. As in the case with $\Delta = 0$, it is observed that the control input remains the same at many instants. Hence,

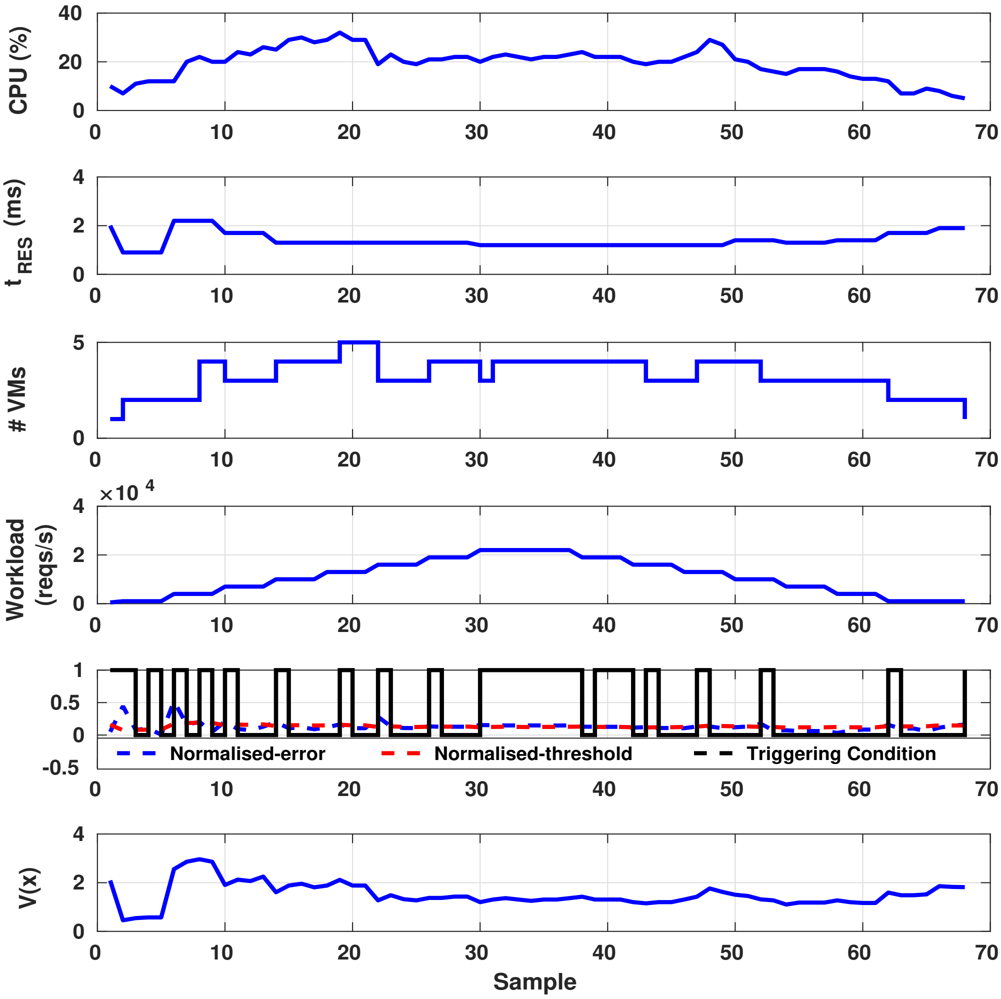

Figure 5.6: Event Triggered Control with theoretical bounds with $\Delta = 2$

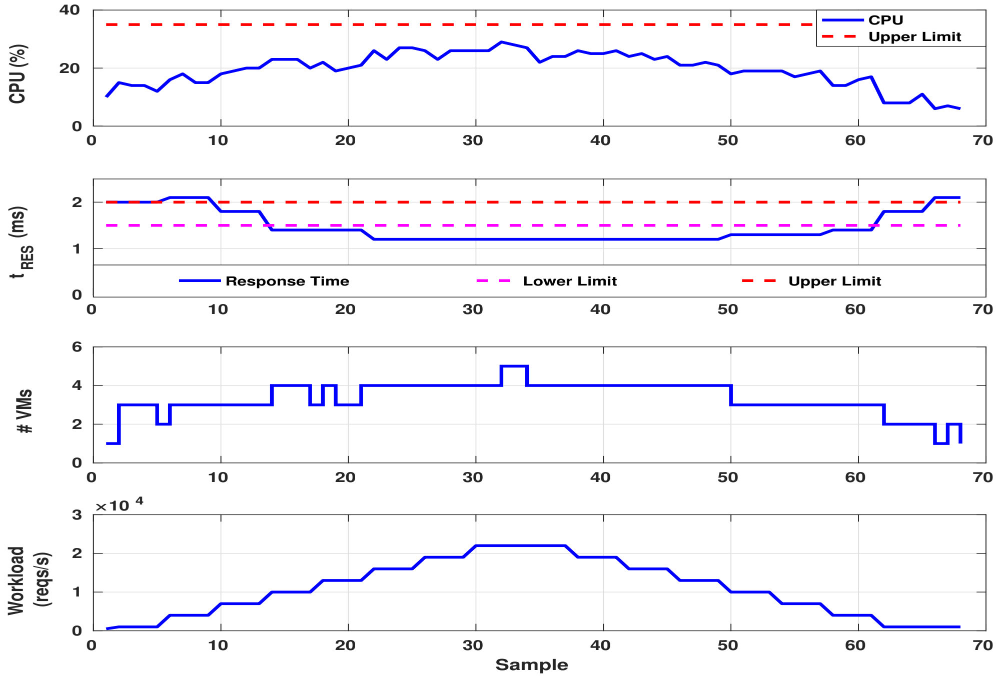

Figure 5.7: Event Triggered Control with practical bounds with $\Delta = 2$

the network bandwidth consumed is high due to frequent updates of control law even though its value is the same over sub-sequent instants. The CPU utilisation increases as the workload increases, and decreases accordingly whilst meeting the performance specifications. The response time decreases in when the workload increases (as in the case with $\Delta = 0$) since there are more number of VMs serving the web-requests.

Figure 5.6 shows the evolution of the system states along with the control input, the workload, the triggering condition and the ISS-Lyapunov function with event triggered control using the triggering condition in (4.7) with $\Delta = 2$. The control law is updated whenever an event occurs, denoted by the triggering condition taking the value '1'. The control input is held constant when the triggering condition is '0'. It is observed that the updates are required less often than in the case of time triggered control, while satisfying performance specifications. Note that the updates are more frequent than in the case of event triggering control with $\Delta = 0$. A decrease in the control input is observed as the workload decreases due to the usage of $\mathcal{D}_R$ regions. Thereby, preventing over-provisioning of resources. This approach results in lower bandwidth consumption as well as efficient resource utilisation.

Figure 5.7 shows the evolution of the system states along with the control input and the workload, with event triggered control using practical bounds on the system states as in (5.2) and (5.3) with $\Delta = 2$. The control law is updated whenever the response time goes outside the desired range or when the CPU utilisation goes above its upper threshold. Similar to the case with event triggered control for $\Delta = 0$, it is observed that while the specification on CPU utilisation is always met, the response time tends to go lower than required due to larger number of VMs than necessary. Whenever such an event occurs, it is seen that the control input tends to decrease in the future instants to bring the response time back within the desired range. The control input is held constant when there is no event. Once again, this approach results in less frequent updates of the control input as compared to time triggered control. As was the case for $\Delta = 0$, the performance of event triggered control with the triggering condition in (4.7) is better than that in this case.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

In this project, an event triggered control strategy is derived for a discrete LTI system subject to exogenous inputs and actuation delay. The concepts of Input to State Stability and $\mathcal{D}_R$ regions are used to arrive at the triggering conditions required for update of the control input. The computational tasks are reduced due to the aperiodic updates of the control input. Event triggered control can become conservative by over-provisioning resources to meet performance specifications. This problem is addressed by choosing poles of the closed loop system from a $\mathcal{D}_R$ region. The designed event triggered control mechanism is validated on an experimental set up. The results demonstrate the ability of event triggered control to ensure efficient physical and computational resource utilization while maintaining quality of service.

Future Work

Some topics of ongoing work include a generalisation of the event triggering condition for general discrete LTI systems by removing the restriction on the norm of the closed loop system matrix $\|\tilde{A}\| < 1$. The derived event triggering condition is to be tested on fast varying workloads in the presence of delay, so as to analyse the maximum delay allowable whilst maintaining system performance. Specific metrics based on system performance and resources consumed are to be formulated to mathematically confirm the savings and efficiency in resource utilisation with event triggering as compared to time triggering. Other models for the web server system such as linear parameter varying models for different operating regions, and non-linear models are to be explored.

REFERENCES

[1] (2017). Haproxy - the reliable, high performance tcp/http load balancer. URL `http://www.haproxy.org/`.

[2] (2017). Kvm. URL `https://www.linux-kvm.org`.

[3] (2017). Matlab - mathworks. URL `https://in.mathworks.com/product/ltc/matlab.html`.

[4] (2018). 57th ieee conference on decision and control. URL `http://cdc2018.ieeecss.org/`.

[5] (2018). Schur complement. URL `http://en.wikipedia.org/wiki/Schur_complement`.

[6] **Boyd, S.**, **L. El Ghaoui**, **E. Feron**, and **V. Balakrishnan**, *Linear matrix inequalities in system and control theory*, volume 15. Siam, 1994.

[7] **Chilali, M.** and **P. Gahinet** (1996). H infin; design with pole placement constraints: an lmi approach. *IEEE Transactions on Automatic Control*, **41**(3), 358–367. ISSN 0018-9286.

[8] **de la Pena, D. M.**, **T. Alamo**, **M. Lazar**, and **W. Heemels** (2008). Design of iss-lyapunov functions for discrete-time linear uncertain systems. *IFAC Proceedings Volumes*, **41**(2), 1135–1140.

[9] **Dimarogonas, D. V.**, **E. Frazzoli**, and **K. H. Johansson** (2012). Distributed event-triggered control for multi-agent systems. *IEEE Transactions on Automatic Control*, **57**(5), 1291–1297.

[10] **Donkers, M.**, **P. Tabuada**, and **W. Heemels** (2014). Minimum attention control for linear systems. *Discrete Event Dynamic Systems*, **24**(2), 199–218.

[11] **Eqtami, A.**, **D. V. Dimarogonas**, and **K. J. Kyriakopoulos**, Event-triggered control for discrete-time systems. *In American Control Conference (ACC), 2010*. IEEE, 2010.

[12] **Heemels, W.**, **K. H. Johansson**, and **P. Tabuada**, An introduction to event-triggered and self-triggered control. *In Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*. IEEE, 2012.

[13] **httperf** (2017). Http performance measurement tool. URL `http://www.hpl.hp.com/research/linux/httperf/httperf-man-0.9.pdf`.

[14] **Kolios, P.**, **C. Panayiotou**, **G. Ellinas**, and **M. Polycarpou** (2016). Data-driven event triggering for iot applications. *IEEE Internet of Things Journal*, **3**(6), 1146–1158.

[15] **Mazo, M.** and **P. Tabuada** (2011). Decentralized event-triggered control over wireless sensor/actuator networks. *IEEE Transactions on Automatic Control*, **56**(10), 2456–2461.

[16] **Peaucelle, D.**, **D. Arzelier**, **O. Bachelier**, and **J. Bernussou** (2000). A new robust d-stability condition for real convex polytopic uncertainty. *Systems & control letters*, **40**(1), 21–30.

[17] **Saikrishna, P. S.**, **R. Pasumarthy**, and **N. P. Bhatt** (2017). Identification and multivariable gain-scheduling control for cloud computing systems. *IEEE Transactions on Control Systems Technology*, **25**(3), 792–807.

[18] **Sontag, E. D.** (1989). Smooth stabilization implies coprime factorization. *IEEE transactions on automatic control*, **34**(4), 435–443.

[19] **Tabuada, P.** (2007). Event-triggered real-time scheduling of stabilizing control tasks. *IEEE Transactions on Automatic Control*, **52**(9), 1680–1685.

LIST OF PAPERS BASED ON THESIS

1. *(submitted)* Kshama Dwarakanath, Durgesh Singh and Ramkrishna Pasumarthy, Event Triggered Control for performance management in web-services in a cloud environment, *Control and Decisions Conference, 2018*.