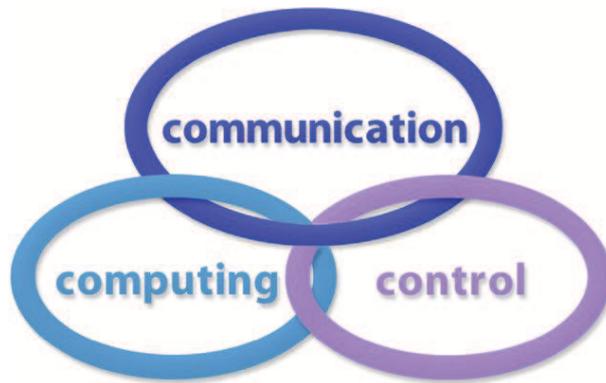


INTERNATIONAL JOURNAL  
of  
COMPUTERS COMMUNICATIONS & CONTROL

ISSN 1841-9836



A Bimonthly Journal  
With Emphasis on the Integration of Three Technologies

Year: 2018 Volume: 13 Issue: 2 Month: April

This journal is a member of, and subscribes to the principles of, the Committee on Publication Ethics (COPE).



<http://univagora.ro/jour/index.php/ijccc/>

**CCC Publications**

Copyright © 2006-2018 by Agora University & CC BY-NC

## BRIEF DESCRIPTION OF JOURNAL

**Publication Name:** International Journal of Computers Communications & Control.

**Acronym:** IJCCC; **Starting year of IJCCC:** 2006.

**ISO:** Int. J. Comput. Commun. Control; **JCR Abbrev:** INT J COMPUT COMMUN.

**International Standard Serial Number:** ISSN 1841-9836.

**Publisher:** CCC Publications - Agora University of Oradea.

**Publication frequency:** Bimonthly: Issue 1 (February); Issue 2 (April); Issue 3 (June); Issue 4 (August); Issue 5 (October); Issue 6 (December).

**Founders of IJCCC:** Ioan DZITAC, Florin Gheorghe FILIP and Misu-Jan MANOLESCU.

### Indexing/Coverage:

- Since 2006, Vol. 1 (S), IJCCC is covered by Thomson Reuters/Clarivate Analytics and is indexed in ISI Web of Science/Knowledge: Science Citation Index Expanded.

2017 Journal Citation Reports® Science Edition (Thomson Reuters, 2016):

*Subject Category:* (1) Automation & Control Systems: Q4(2009, 2011, 2012, 2013, 2014, 2015), Q3(2010, 2016); (2) Computer Science, Information Systems: Q4(2009, 2010, 2011, 2012, 2015), Q3(2013, 2014, 2016).

Impact Factor/3 years in JCR: 0.373(2009), 0.650 (2010), 0.438(2011); 0.441(2012), 0.694(2013), 0.746(2014), 0.627(2015), **1.374(2016)**.

Impact Factor/5 years in JCR: 0.436(2012), 0.622(2013), 0.739(2014), 0.635(2015), **1.193(2016)**.

- Since 2008 IJCCC is indexed by Scopus: **CiteScore 2016 = 1.06**.

*Subject Category:*

(1) Computational Theory and Mathematics: Q4(2009, 2010, 2012, 2015), Q3(2011, 2013, 2014, 2016);

(2) Computer Networks and Communications: Q4(2009), Q3(2010, 2012, 2013, 2015), Q2(2011, 2014, 2016);

(3) Computer Science Applications: Q4(2009), Q3(2010, 2011, 2012, 2013, 2014, 2015, 2016).

SJR: 0.178(2009), 0.339(2010), 0.369(2011), 0.292(2012), 0.378(2013), 0.420(2014), 0.263(2015), 0.319(2016).

- Since 2007, 2(1), IJCCC is indexed in EBSCO.

**Focus & Scope:** International Journal of Computers Communications & Control is directed to the international communities of scientific researchers in computers, communications and control, from the universities, research units and industry. To differentiate from other similar journals, the editorial policy of IJCCC encourages the submission of original scientific papers that focus on the integration of the 3 "C" (Computing, Communications, Control).

In particular, the following topics are expected to be addressed by authors:

(1) Integrated solutions in computer-based control and communications;

(2) Computational intelligence methods & Soft computing (with particular emphasis on fuzzy logic-based methods, computing with words, ANN, evolutionary computing, collective/swarm intelligence);

(3) Advanced decision support systems (with particular emphasis on the usage of combined solvers and/or web technologies).

## FAMOUS FORMER MEMBER IN THE EDITORIAL BOARD OF IJCCC



Lotfi A. Zadeh (Feb. 4, 1921 - Sept. 6, 2017)  
The inventor of Fuzzy Sets, Fuzzy Logic, and Soft Computing  
Former member in the Editorial Board of IJCCC between 2008-2017

## EXCELLENCE AWARD PROPOSED FOR IJCCC



Journals Excellence Awards proposed by Scopus (2015)

## Prof. Ioan Dzitac at 65 Years



Prof. Ioan Dzitac, PhD,  
Founder and Editor in Chief  
of the

International Journal of Computers Communications & Control.

Ioan Dzitac, born on February 14, 1953, Poienile de sub Munte, Maramures, Romania, received his B.Sc.(eq.M.Sc.) in Mathematics (1977) and, Ph.D. in Information Sciences (2002), from Babes-Bolyai University of Cluj - Napoca, Romania.

He is a Full Professor at Aurel Vlaicu University of Arad (since 2009) and at Agora University of Oradea (since 2017). He was an Adjunct Professor at University of Chinese Academy of Sciences - Beijing, China (2013-2016) and since 2016 is an Advisory Board Member at Graduate School of Management of Technology, Hoseo University, South Korea. He was in managerial positions, such as: Director of Mathematics & Informatics Department (2004-2005), Head of IT Chair (2005-2008), Director of R&D Center (2009-2017). He was Rector of Agora University of Oradea in the period 2012-2016 and, was re-elected for a new mandate, 2016-2020.

He is founder and General Chair of International Conference on Computers Communications and Control and he was member in the Program Committee of more than 80 international conferences. He is a Senior Member of IEEE (since 2011) and other professional organizations (ROMAI, SRAIT, SSMAR, RNG-ISCB).

He was an invited speaker and/or invited special sessions' organizer and chair in China (2013: Beijing, Suzhou and Chengdu, 2015: Dalian, 2016: Beijing), India (2014: Madurai, 2017: Delhi), Russia (2014: Moscow), Brazil (2015: Rio), Lithuania (2015: Druskininkai), South Korea (2016: Asan).

His current research interests include different aspects of artificial intelligence, soft computing and applications of fuzzy logic in technology and economy. He has published 3 books, 12 courses and materials for students, 5 conference proceedings and more than 80 scientific papers in journals and conferences proceedings.

Prof. Ioan Dzitac is included in the volume "100 Romanian Theoretical Computer Scientists", edited by the Romanian Academy (Eds. S. Cojocaru, G. Paun and D. Vaida), a top 100 for the last 100 years (Romania and Republic of Moldova, 1918-2018).

## EDITORIAL STAFF OF IJCCC (2018)

### EDITORS-IN-CHIEF:

#### **Ioan DZITAC**

Aurel Vlaicu University of Arad, Romania  
St. Elena Dragoi, 2, 310330 Arad  
professor.ioan.dzitac@ieee.org

#### **Florin Gheorghe FILIP**

Romanian Academy, Romania  
125, Calea Victoriei, 010071 Bucharest  
fflip@acad.ro

### MANAGING EDITOR:

#### **Mișu-Jan MANOLESCU**

Agora University of Oradea, Romania  
Piata Tineretului, 8, 410526 Oradea  
mmj@univagora.ro

### EXECUTIVE EDITOR:

#### **Răzvan ANDONIE**

Central Washington University, USA  
400 East University Way, Ellensburg, WA 98926  
andonie@cwu.edu

### PROOFREADING EDITOR:

#### **Răzvan MEZEI**

Lenoir-Rhyne University, USA  
Madison, WI  
proof.editor@univagora.ro

### LAYOUT EDITOR:

#### **Horea OROS**

University of Oradea, Romania  
St. Universitatii 1, 410087, Oradea  
horos@uoradea.ro

### TECHNICAL EDITOR:

#### **Domnica Ioana DZITAC**

New York University Abu Dhabi, UAE  
Saadiyat Marina District, Abu Dhabi  
domnica.dzitac@nyu.edu

### EDITORIAL ADDRESS:

Agora University, Cercetare Dezvoltare Agora, Tineretului 8, 410526 Oradea, Bihor, Romania,  
Tel./ Fax: +40 359101032, E-mail: ijccc@univagora.ro, rd.agora@univagora.ro  
URL: <http://univagora.ro/jour/index.php/ijccc/>

## EDITORIAL BOARD OF IJCCC (MEMBERS, 2018):

### **Vandana AHUJA**

Jaypee Institute of Information Technology,  
INDIA  
A-10, Sector-62, Noida 201307, Delhi  
vandana.ahuja@jiit.ac.in

### **Fuad ALESKEROV**

Russian Academy of Sciences, RUSSIA  
HSE, Shabolovka St, Moscow  
alesk@hse.ru

### **Luiz F. AUTRAN GOMES**

Ibmec, Rio de Janeiro, BRAZIL  
Av. Presidente Wilson, 118  
autran@ibmecrj.br

### **Barnabas BEDE**

DigiPen Institute of Technology, USA  
Redmond, Washington  
bbede@digipen.edu

### **Dan BENTA**

Agora University of Oradea, ROMANIA  
Tineretului, 8, 410526 Oradea  
dan.benta@univagora.ro

### **Pierre BORNE**

Ecole Centrale de Lille, FRANCE  
Villeneuve d'Ascq Cedex, F 59651  
p.borne@ec-lille.fr

### **Alfred M. BRUCKSTEIN**

Ollendorff Chair in Science, ISRAEL  
Technion, Haifa 32000  
freddy@cs.technion.ac.il

### **Ioan BUCIU**

University of Oradea, ROMANIA  
Universitatii, 1, Oradea  
ibuciu@uoradea.ro

### **Hariton-Nicolae COSTIN**

Univ. of Med. and Pharmacy, ROMANIA  
St. Universitatii No.16, 6600 Iasi  
hcostin@iit.tuiasi.ro

### **Felisa CORDOVA**

University Finis Terrae, CHILE  
Av. P. de Valdivia 1509, Providencia  
fcordova@uft.cl

### **Petre DINI**

Concordia University, CANADA  
Montreal, Canada  
pdini@cisco.com

### **Antonio Di NOLA**

University of Salerno, ITALY  
Via Ponte Don Melillo, 84084 Fisciano  
dinola@cds.unina.it

### **Yezid DONOSO**

Univ. de los Andes, COLOMBIA  
Cra. 1 Este No. 19A-40, Bogota  
ydonoso@uniandes.edu.co

### **Gintautas DZEMYDA**

Vilnius University, LITHUANIA  
4 Akademijos, Vilnius, LT-08663  
gintautas.dzemyda@mii.vu.lt

### **Simona DZITAC**

University of Oradea, ROMANIA  
1 Universitatii, Oradea  
simona@dzitac.ro

### **Ömer EGEÇIOĞLU**

University of California, USA  
Santa Barbara, CA 93106-5110  
omer@cs.ucsb.edu

### **Constantin GAINDRIC**

IMMAS, Republic of MOLDOVA  
Kishinev, 277028, Academiei 5  
gaindric@math.md

### **Xiao-Shan GAO**

Academia Sinica, CHINA  
Acad. of Math. and System Sci.  
Beijing 100080, China  
xgao@mmrc.iss.ac.cn

**Enrique HERRERA-VIEDMA**  
University of Granada, SPAIN  
Av. del Hospicio, s/n, 18010 Granada  
viedma@decsai.ugr.es

**Kaoru HIROTA**  
Tokyo Institute of Tech., JAPAN  
G3-49,4259 Nagatsuta  
hirota@hrt.dis.titech.ac.jp

**Gang KOU**  
SWUFE, CHINA  
Chengdu, 611130  
kougang@swufe.edu.cn

**Heeseok LEE**  
KAIST, KOREA  
85 Hoegiro, Seoul 02455  
hsl@business.kaist.ac.kr

**George METAKIDES**  
University of Patras, GREECE  
Patra 265 04, Greece  
george@metakides.net

**Shimon Y. NOF**  
Purdue University, USA  
610 Purdue Mall, West Lafayette,  
IN 47907, USA  
nof@purdue.edu

**Stephan OLARIU**  
Old Dominion University, USA  
Norfolk, VA 23529-0162  
olariu@cs.odu.edu

**Gheorghe PĂUN**  
Romanian Academy, ROMANIA  
IMAR, Bucharest, PO Box 1-764  
gpaun@us.es

**Mario de J. PEREZ JIMENEZ**  
University of Seville, SPAIN  
Avda. Reina Mercedes s/n, 41012  
marper@us.es

**Radu-Emil PRECUP**  
Pol. Univ. of Timisoara, ROMANIA  
Bd. V. Parvan 2, 300223  
radu.precup@aut.upt.ro

**Radu POPESCU-ZELETIN**  
Technical University Berlin, GERMANY  
Fraunhofer Institute for Open CS  
rpz@cs.tu-berlin.de

**Imre J. RUDAS**  
Obuda University, HUNGARY  
Budapest, Becsı ut 96b, 1034  
rudas@bmf.hu

**Yong SHI**  
Chinese Academy of Sciences, CHINA  
Beijing 100190  
yshi@gucas.ac.cn, yshi@unomaha.edu

**Bogdana STANOJEVIC**  
Serbian Academy of SA, SERBIA  
Mathematical Institute  
Kneza Mihaila 36, Beograd 11001  
bgdnpop@mi.sanu.ac.rs

**Athanasios D. STYLIADIS**  
University of Kavala, GREECE  
65404 Kavala  
styliadis@teikav.edu.gr

**Gheorghe TECUCI**  
George Mason University, USA  
University Drive 4440, Fairfax VA  
tecuci@gmu.edu

**Horia-Nicolai TEODORESCU**  
Romanian Academy, ROMANIA  
Iasi Branch, Bd. Carol I 11, 700506  
hteodor@etc.tuiasi.ro

**Dan TUFIS**  
Romanian Academy, ROMANIA  
Research Institute for AI  
13 Septembrie, 13, 050711 Bucharest  
tufis@racai.ro

## Contents

<b>Adaptive-Smith Predictor for Controlling an Automotive Electronic Throttle over Network</b>	
C.F. Caruntu, A.N. Vargas, L. Acho, G. Pujol	151
<b>Moving Object Detection and Tracking using Genetic Algorithm Enabled Extreme Learning Machine</b>	
G. Jemilda, S. Baulkani	162
<b>A Hybrid Failure Diagnosis and Prediction using Natural Language-based Process Map and Rule-based Expert System</b>	
D. Kim, Y. Lin, S. Lee, B.H. Kang, S.C. Han	175
<b>Decoupling 5G Network Control: Centralised Coordination and Distributed Adap- tation</b>	
S.J. Lin, J.G. Yu, W. Ni, R.P. Liu	192
<b>New Failure Mode and Effects Analysis based on D Numbers Downscaling Method</b>	
B. Liu, Y. Hu, Y. Deng	205
<b>An Improved Local Descriptor based Object Recognition in Cluttered 3D Point Clouds</b>	
X. Liu, Y. Lu, T. Wu, T. Yuan	221
<b>Estimating Warehouse Rental Price using Machine Learning Techniques</b>	
Y. Ma, Z. Zhang, A. Ihler, B. Pan	235
<b>Control-Scheduling Codesign for NCS based Fuzzy Systems</b>	
P.E. Mendez-Monroy, I. Sanchez Dominguez, A. Bassam, O. May Tzuc	251
<b>EPAK: A Computational Intelligence Model for 2-level Prediction of Stock In- dices</b>	
L. Tang, H. Pan, Y. Yao	268
<b>Outlook of Coordinated Transmission Control in 5G Networks for IoTs</b>	
U. Tariq, A. Aldaej	280
<b>Author index</b>	294

# Adaptive-Smith Predictor for Controlling an Automotive Electronic Throttle over Network

C.F. Caruntu, A.N. Vargas, L. Acho, G. Pujol

## Constantin F. Caruntu

Gheorghe Asachi Technical University of Iasi,  
Str. Prof. Dimitrie Mangeron, 27, 700050, Iasi, Romania  
caruntuc@ac.tuiasi.ro

## Alessandro N. Vargas\*

Universidade Tecnológica Federal do Paraná, UTFPR,  
Av. Alberto Carazzai 1640,  
86300-000 Cornélio Procópio-PR, Brazil  
\*Corresponding author: avargas@utfpr.edu.br

## Leonardo Acho and Gisela Pujol

Universitat Politècnica de Catalunya Barcelona Tech,  
Escola d'Enginyeria de Barcelona Est,  
CoDALab (Control, Dynamics and Applications),  
Carrer d'Eduard Maristany, 10-14, 08930 Barcelona, Spain  
leonardo.acho@upc.edu, gisela.pujol@upc.edu

**Abstract:** The paper presents a control strategy for an automotive electronic throttle, a device used to regulate the power produced by spark-ignition engines. Controlling the electronic throttle body is a difficult task because the throttle accounts strong nonlinearities. The difficulty increases when the control works through communication networks subject to random delay. In this paper, we revisit the Smith-predictor control, and show how to adapt it for controlling the electronic throttle body over a delay-driven network. Experiments were carried out in a laboratory, and the corresponding data indicate the benefits of our approach for applications.

**Keywords:** Adaptive-Smith predictor, electronic throttle control, networked control systems, switching control.

## 1 Introduction

Electronic throttle control represents a novel technology that has been used over the last years by the automotive industry to improve the efficiency of spark-ignition engines [6–8, 12, 14, 20, 24, 26]. The electronic throttle control system comprises four parts: (i) accelerator pedal module, (ii) engine/transmission control module, (iii) electronic throttle body (ETB), and (iv) network.

This paper focus on controlling the electronic throttle body (ETB) through a network.

Designing a controller for the ETB is a difficult task because the ETB has a nonlinear behavior—nonlinearities arise from gear backlash, friction on the mechanical components, limp-home effects, and nonlinear forces provoked by the security return spring [8, 21–23, 26, 28–30].

To control the ETB, the current literature proposes different approaches, such as PID controller with system nonlinearities compensated by a heuristically tuned compensator [6], dynamic programming techniques [9], variable-structure control [20], robust discrete-time model reference adaptive control [7], [14], PID controller with adaptive compensator for friction, limp-home, and backlash [8], mixed constrained  $H_2/H_\infty$  controller [26], sliding-mode controller [12], asymmetric nonlinear PI controller [24], just to name a few.

The aforementioned control strategies require that sensors, controllers, and actuators are directly connected by wires, an unfeasible requirement in many cases—in modern vehicles, for

instance, control signals from controllers and measurements from sensors are exchanged using a communication network, e.g., Controller Area Network (CAN), see [11, 17].

Communication through network brings up a challenge on how to handle the effects of the network-induced delays in the control loop. Delays are time-varying by its own nature—random driven—and they degrade the performance of control systems—recall that delays can destabilize closed-loop systems, see [3, 4, 13].

Smith-predictor control appeared as a seminal approach for controlling systems subject to fixed (deterministic) delay [18, Ch. 5]. Afterwards in the literature, the classical Smith predictor was improved to deal with time-varying delays, a strategy referred to as *adaptive* Smith predictor (ASP), showing promising results [5, 10, 25]. This paper shows an extension, a contribution towards the application of ASP in the control of automotive electronic throttle body.

Not only does the contribution of the paper rely on applying the ASP in the control of the throttle, but also it does present an improved ASP, a novel strategy that accounts the asymmetry of the throttle's valve. This asymmetry was firstly observed in [24]. Here, we modify the ASP strategy to account this asymmetry—checking this asymmetric-ASP strategy in practice sets the main contribution of this paper.

Our approach has implications for the automotive industry. Indeed, experiments were carried out in a laboratory testbed in which an electronic throttle body was controlled; the corresponding outcome illustrates the benefits of our approach.

## 2 Adaptive Smith predictor

In [10], the authors suggest using the classical Smith predictor to control systems subject to constant delay. However, the results of [10] do not apply when the underlying communication network is either random or time-varying.

Our main contribution relies upon adapting the Smith-predictor strategy to control an electronic throttle when the communication network has random delays. Our adapted strategy first recalls the so-called *adaptive Smith predictor* (ASP) from [10]. Recall that ASP stands for a scheme that modifies the classical Smith predictor so as to embed it with information of the network-induced time-varying delay. Second, we modify the ASP to include the asymmetry observed in [24], as detailed next.

Fig. 1 shows the ASP scheme. It also shows the process to be controlled, i.e.,  $G_p(s)$ , as well as a communication network with delays in the forward path— $\tau^{ca}$ —and in the feedback path— $\tau^{sc}$ . The ASP scheme is composed by a controller  $C(s)$ , a process model  $G_m(s)$ , and a delay term that accounts the delays measured in the network  $\tau^{est}$ .

As usual,  $y(t)$  represents the output,  $u(t)$  denotes the control input, and  $r(t)$  represents the reference signal. The error signal,  $e(t) = y(t) - r(t)$ , should be made as small as possible.

The subscripts  $s$  and  $r$  in the variables denote ‘sent’ and ‘received’ signals, respectively, which allow us to write

$$u_r(t) = u_s(t - \tau^{ca}(t)) \quad \text{and} \quad y_r(t) = y_s(t - \tau^{sc}(t)). \quad (1)$$

### 2.1 Closed-loop analysis

For sake of notational simplicity, let us assume that the delays from the network links, i.e.,  $\tau^{ca}$  and  $\tau^{sc}$ , are time invariant. In this case, the closed-loop transfer function of the block-oriented structure of Fig. 1 reads as

$$\frac{Y_s(s)}{R(s)} = \frac{C(s)G_p(s)e^{-\tau^{ca}s}}{1 + C(s)G_m(s) + N(s)}, \quad (2)$$

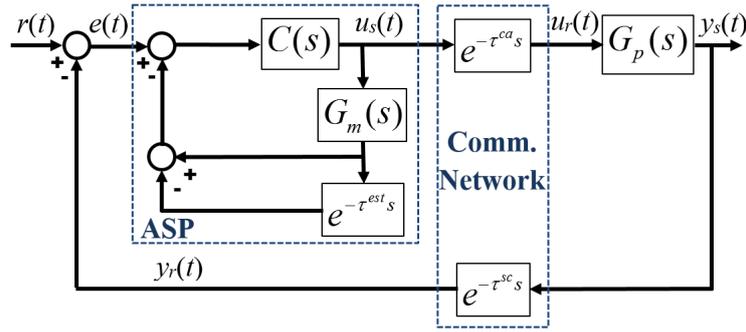


Figure 1: Adaptive Smith-predictor (ASP) controller in a closed-loop diagram. Controller communicates with the plant through a network.

where

$$N(s) = C(s) \left[ G_p(s) e^{-(\tau^{ca} + \tau^{sc})s} - G_m(s) e^{-\tau^{est}s} \right]. \quad (3)$$

*Remark 1.* When  $\tau^{ca}(t)$  and  $\tau^{sc}(t)$  are time varying, one can employ the linear state-space approach to represent the block-oriented structure of Fig. 1, an approach equivalent to (2) (see [19, Sec. 2]).

The delays  $\tau^{ca}$  and  $\tau^{sc}$  can dramatically degrade the stability of the system—we want to minimize the negative effects of such delays in the stability. To do so, we need

$$N(s) \approx 0. \quad (4)$$

The expression in (4) does hold provided that not only the prediction model approximates the plant process, i.e.,  $G_m(s) \approx G_p(s)$ , but also the estimated delay approximates the actual network-induced time-varying delay, i.e.,  $\tau^{est} \approx \tau^{sc} + \tau^{ca}$ . In this case, the condition in (4) holds true so that (2) equals

$$\frac{Y_s(s)}{R(s)} = \frac{C(s)G_m(s)e^{-\tau^{ca}s}}{1 + C(s)G_m(s)}. \quad (5)$$

A benefit drawn from the expression in (5) when compared to (2) is that delays cannot destabilize the system in (5).

Hereafter, we consider two assumptions: (i) the model  $G_m(s)$  matches the process  $G_p(s)$ ; and (ii) the estimated delay  $\tau^{est}$  approximates the total communication delay  $\tau^{sc} + \tau^{ca}$  (see [15, 16]).

## 2.2 Switching PID controller

In [24], the authors have observed that the behavior of the throttle valve depends on whether it is opening or closing, a feature called *asymmetric performance*. This feature is considered into the design of the adaptive Smith predictor based on a switching PID control, as follows.

The opening and closing behavior of the throttle followed the next switching rule:

$$\begin{aligned} \dot{y}_r(t) &\geq 0: \text{ opening phase } (j = 1), \\ \dot{y}_r(t) &< 0: \text{ closing phase } (j = 2). \end{aligned} \quad (6)$$

Then the switching-PID controller (SPID) reads as

$$C_j(s) = K_{p,j} \left( 1 + \frac{1}{T_{i,j}s} + \frac{T_{d,j}s}{1 + \alpha_j s} \right), \quad j = 1, 2, \quad (7)$$

where the constants  $K_p$ ,  $T_i$ , and  $T_d$  represent the proportional, integrative, and derivative elements, respectively.

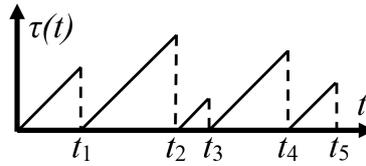


Figure 2: Example of the delays describing the zero-order hold function on an aperiodic realization.

### Modeling the electronic throttle valve

The literature acknowledges that nonlinear systems are the most appropriate ones to represent the electronic throttle valve [8, 21–23, 26, 28–30]; however, for sake of simplicity, we decided to let it be represented by a first order process with dead-time in the form

$$G_m(s) = \frac{K}{Ts + 1} e^{-Ls}. \quad (8)$$

This simplification was purposeful to alleviate the underlying numerical burden of real-time implementation (see Section 3).

*Remark.* The *asymmetric* switching-PID control represents the scheme of Fig. 1, together with (7), and also with (8) modified to consider asymmetry, i.e.,  $G_{m,j}(s)$ ,  $j = 1, 2$ .

### 2.3 Delay in the communication channel modeled as a Poisson process

Recall the scheme of Fig. 1. By assumption, both delays on the forward channel  $\tau^{ca}(t)$  and the feedback channel  $\tau^{sc}(t)$  assume a sawtooth format, as illustrated in Fig. 2.

In addition, both delays satisfy  $\dot{\tau}(t) = 1$ , for all  $t > 0$ , almost everywhere, with discontinuous reset-to-zero points occurring at instants  $t_0 = 0 < t_1 < t_2 < \dots$ , which correspond to the arrival times of a homogeneous Poisson process with rate  $\lambda$ . The inter-arrival times  $\tau_k := t_{k+1} - t_k$ ,  $k \geq 0$ , form independent and identically distributed times, and the corresponding probability density function of the inter-arrival time  $\tau_k$  reads as

$$\Pr[\tau_k = t] = \lambda e^{-\lambda t}, \quad \forall t \geq 0. \quad (9)$$

### 2.4 Delay in the ASP strategy

The adaptive Smith-predictor (ASP) strategy in Fig. 1 requires the value  $\tau^{est}$ —recall that it should approximate the total communication delay  $\tau^{sc} + \tau^{ca}$ , as mentioned in Section 2.1.

At each sampling step, say  $k > 0$ , we measure the delay from the  $(k - 1)$ -th step using a time-stamping technique. Next, following the suggestion of [2], we replace the term  $\tau^{est}$  in the ASP scheme (see Fig. 1) by the time-varying process

$$\tau_k^{est} = \frac{2}{N} \sum_{\ell=k-N}^{k-1} \tau_\ell^{sc}, \quad (10)$$

where  $\tau_\ell^{sc}$  represents the communication delay from sensor-to-controller. The constant "2" in (10) arises from the simplifying assumption that  $\tau_\ell^{sc}$  and  $\tau_\ell^{ca}$  are identical.

## 3 Experimental results

This section presents experimental data obtained from a laboratory testbed, designed to control an electronic throttle body through a communication-emulated network.



Figure 3: Laboratory testbed. Experimental setup for controlling the electronic throttle: Quanser power amplifier, Quanser real-time control board, and ETB.

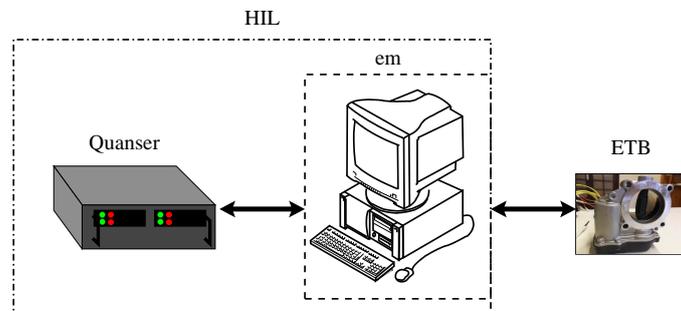


Figure 4: Diagram representing the setup assembled in the laboratory testbed. The Hardware-in-the-loop (HIL) device sends and receives signals from the electronic throttle body (ETB), and the signals reaches the Quanser board through a computer that emulated a network

### 3.1 Hardware-in-the-loop platform

The control strategy of Fig. 1 was implemented in a laboratory testbed that included both an automotive electronic throttle and a Hardware-in-the-Loop (HIL) device (Fig. 3). The laboratory was composed by the following elements [24]:

- Quanser Q4 real-time control board that sets the operation clock of the closed-loop (effective control, discrete time) at a fixed sampling rate of  $T_s = 1$  ms;
- Quanser UPM180-25-B-PWM power amplifier (to supply the voltage and electrical current consumed by the equipment);
- Continental VDO electronic throttle body (c.f, [24]). The throttle device is equipped with a position sensor, an important device in feedback control systems [27]. The position sensor generates a proportional voltage ranging from 0 V (completely closed) to +5 V (completely opened);
- Computer that emulated a communication network with random delays, following the scheme of Fig. 4—delays were driven by a Poisson process (see Section 2.3).

### 3.2 Control methods: experimental comparison

For comparison purposes, six controllers were designed and checked in the laboratory testbed, all under the scheme of Fig. 1, as follows:

1. PID controller in  $C(s)$ . Set  $G_m(s) \equiv 0$  and  $\tau^{est} \equiv 0$ ;
2. Switching PID controller (SPID) as in (7). Set  $G_m(s) \equiv 0$  and  $\tau^{est} \equiv 0$ ;
3. Classical Smith Predictor based on the PID controller (SP-PID);
4. Classical Smith Predictor based on the switching PID controller (SP-SPID);
5. Adaptive Smith Predictor based on the PID controller (ASP-PID);
6. Adaptive Smith Predictor based on the switching PID controller (ASP-SPID).

The experiments that were carried out in the laboratory considered the aforementioned six control strategies with parameters designed by the Approximate M-constrained Integral Gain Optimization (AMIGO) tuning rules [1], which were calculated for PID tuning in process control:

$$\begin{aligned}
 K_p &= \frac{1}{K} \left( 0.2 + 0.45 \frac{T}{L} \right) - \text{proportional term,} \\
 T_i &= \frac{0.4L + 0.8T}{L + 0.1T} L - \text{integral term,} \\
 T_d &= \frac{0.5LT}{0.3L + T} - \text{derivative term.}
 \end{aligned} \tag{11}$$

The switching strategies, in particular, were designed by setting a simple average model  $(G_{m,1} + G_{m,2})/2$  into AMIGO with  $T = 0.4$ .

#### First-order representation

Aiming for simplicity in the computational implementation, we decided to represent  $G_m(s)$  as a first-order linear model (e.g., Section 2.2). Step responses allowed us to obtain

$$G_{m,j}(s) = \frac{19.72}{a_j s + 1} e^{-0.025s}, \quad j = 1, 2, \tag{12}$$

where  $j = 1, 2$  corresponds to the opening and closing valve, namely,  $a_1 = 0.65$  (opening) and  $a_2 = 0.15$  (closing).

### 3.3 Experimental results

Experiments were carefully performed to control the electronic throttle under the six proposed control strategies.

The communication-emulated network were checked under two Poisson processes: high load ( $\lambda = 100$ ) and low load ( $\lambda = 250$ ). A sample of Poisson delays used in the experiments is illustrated in Fig. 5 and 6. This sample was kept fixed for all control methods, just for sake of a fair comparison.

The reference input signal was constructed as follows. Two sinusoidal signals were summed up, one signal with amplitude of 25 degrees centered around 45 degrees and frequency of 1 rad/s, and the other with amplitude of 3 degrees and frequency of 5 rad/s. This reference signal was

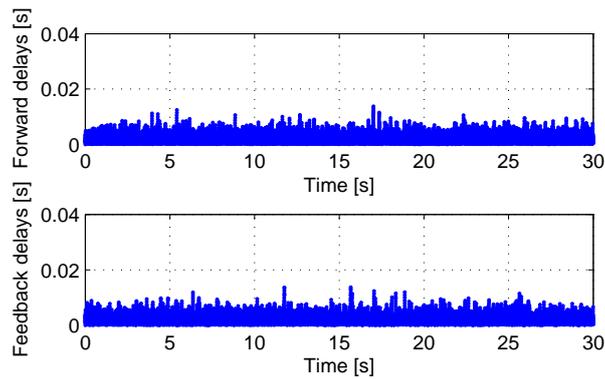


Figure 5: Part A ( $\lambda = 250$ ): Forward channel and feedback channel delays.

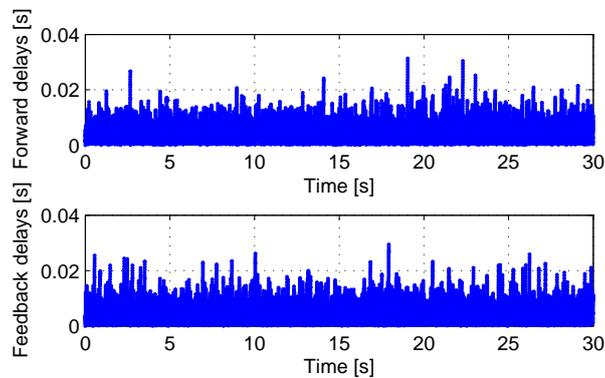


Figure 6: Part B ( $\lambda = 100$ ): Forward channel and feedback channel delays.

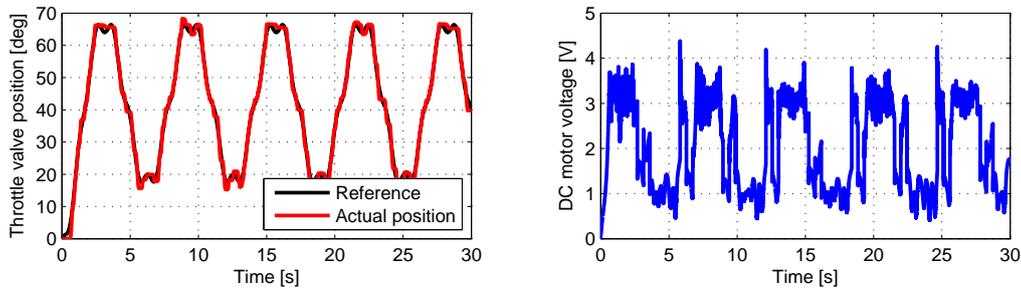


Figure 7: Part A ( $\lambda = 250$ ): Curves for the ASP-SPID control. Throttle valve position.

useful to analyze how the controllers actuated not only to large and small variations, but also to slow and fast variations of the throttle valve position.

Experimental outcome obtained with the ASP-SPID control architecture for the two different network loads is summarized in Figs. 7 and 8. As can be seen, the effect of time-varying delays introduced by the communication network on the closed-loop seems negligible.

Figs. 9 and 10 show the corresponding experimental error signal. Data suggest that, even though the error increased with the augmented load of the network, the error kept bounded by 7 degrees, an acceptable value according to [14].

Evaluation of the classical PID was also considered in the laboratory—not surprisingly, the classical PID showed an unstable behavior. This finding emphasizes the usefulness of our approach for controlling the electronic throttle through networks subject to random delays.

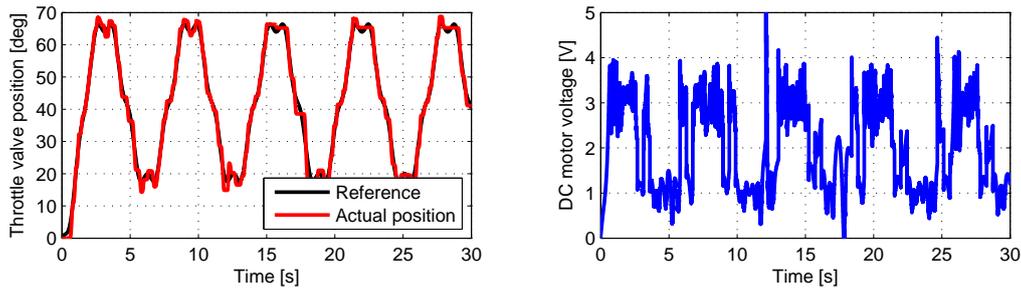


Figure 8: Part B ( $\lambda = 100$ ): Curves for the ASP-SPID control. Throttle valve position.

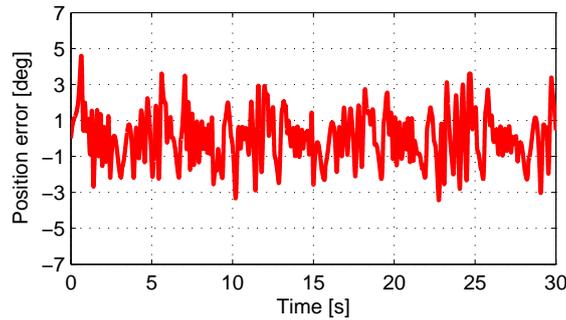


Figure 9: Part A ( $\lambda = 250$ ): Curves for the ASP-SPID control. Error signal.

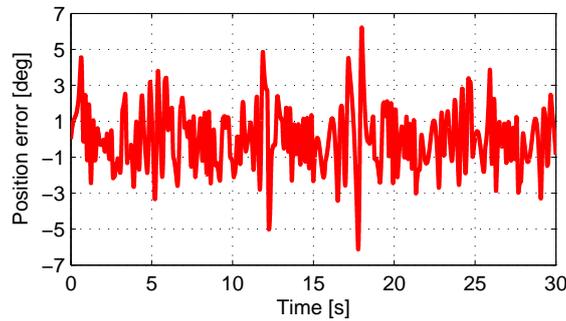


Figure 10: Part B ( $\lambda = 100$ ): Curves for the ASP-SPID control. Error signal.

Finally, to complete the experimental analysis, we computed the cumulative error  $I = \sum_{k=0}^{\infty} e_k^2$  for all methods— $e_k$  represents the experimental, measured error at the  $k$ -th sampling time, see Table 1. As can be seen, our switching strategy—ASP-SPID—produced the lowest error for both low and high loads. This experimental evidence suggests that the switching control be a useful tool for real-time applications.

Table 1: Experimental error index for an automotive throttle body.

Controller	Low load	High load
PID	$\infty$	$\infty$
SPID	$\infty$	$\infty$
SP-PID	79.61	111.02
SP-SPID	65.81	84.20
ASP-PID	64.41	80.62
ASP-SPID	55.37	76.66

## 4 Concluding remarks

The paper revisits the adaptive Smith-predictor control, and shows how modify it to incorporate the asymmetry of the throttle—a feature observed in [24]. The asymmetry generated a switching PID-based control, referred to as *Adaptive Smith Predictor based on the switching PID controller*. This controller was checked in a real-time experiment, that for controlling an automotive electronic throttle device over a random-delay-driven network—the proposed controller showed promising experimental results.

## Acknowledgements

Research supported in part by the Spanish Ministry of Economy and Competitiveness through the research projects DPI2015-64170-R/MINECO/FEDER, DPI2011-25822; by the Government of Catalonia (Spain) through 2014SGR859; and by the Brazilian agencies CNPq Grant 401572/2016-1; 307882/2014-4; and CAPES Grant Programa PVE 88881.030423/2013-01. This work was partially supported by the Spanish Ministry of Economy, Industry and Competitiveness, under grants DPI2016-77407-P (AEI/FEDER, UE).

## Bibliography

- [1] Astrom, K.J.; Haggglund, T. (2004); Revisiting the Ziegler-Nichols step response method for PID control, *Journal of Process Control*, 14, 635–650, 2004.
- [2] Balau, A.E.; Caruntu, C.F.; Lazar, C.(2011); Simulation and control of an electro-hydraulic actuated clutch, *Mechanical Systems and Signal Processing*, 25, 1911–1922, 2011.
- [3] Caruntu, C.F.; Lazar, C. (2014); Network delay predictive compensation based on time-delay modeling as disturbance. *International Journal of Control*, 87, 2012–2026, 2014.
- [4] Caruntu, C.F.; Lazar, M.; Gielen, R.H.; van den Bosch, P.P.J.; Di Cairano, S. (2013); Lyapunov based predictive control of vehicle drivetrains over CAN, *Control Engineering Practice*, 21, 1884–1898, 2013.
- [5] Chen, C.H.; Lin, C.L.; Hwang, T.S.(2007); Stability of networked control systems with time-varying delays, *IEEE Communication Letters*, 11, 270–272, 2007.
- [6] Deur, J.; Pavkovic, D.; Peric, N.; Jansz, M.; Hrovat, D.(2004); An electronic throttle control strategy including compensation of friction and limp-home effects, *IEEE Transactions on Industry Applications*, 40, 821–834, 2004.
- [7] di Bernardo, M.; di Gaeta, A.; Montanaro, U.; Olm, J.M.; Santini, S.(2013); Experimental validation of the discrete-time MCS adaptive strategy, *Control Engineering Practice*, 21, 847–859, 2013.
- [8] Jiao, X.; Zhang, J.; Shen, T. (2014); An adaptive servo control strategy for automotive electronic throttle and experimental validation, *IEEE Transactions on Industrial Electronics*, 61, 6275–6284, 2014.
- [9] Kim, D.; Peng, H.; Bai, S.; Maguire, J.M. (2007); Control of integrated powertrain with electronic throttle and automatic transmission, *IEEE Transactions on Control Systems Technology*, 15, 474–482, 2007.

- 
- [10] Lai, C.L.; Hsu, P.L. (2010); Design the remote control system with the time-delay estimator and the adaptive Smith predictor, *IEEE Transactions on Industrial Informatics*, 6, 73–80, 2010.
- [11] Lee, K.C.; Kim, M.H.; Lee, S.; Lee, H.H. (2004); IEEE-1451-Based smart module for in-vehicle networking systems of intelligent vehicles, *IEEE Transactions on Industrial Electronics*, 51, 1150–1158, 2004.
- [12] Li, Y.; Yang, B.; Zheng, T.; Li, Y.; Cui, M.; Peeta, S. (2015); Extended-state-observer-based double-loop integral sliding-mode control of *IEEE Transactions on Intelligent Transportation Systems*, 16, 2501–2510, 2015.
- [13] Mahmoud, M.S.(2014); *Control and Estimation Methods over Communication Networks*, Springer International Publishing Switzerland, 2014.
- [14] Montanaro, U.; di Gaeta, A.; Giglio, V. (2014); Robust discrete-time MRAC with minimal controller synthesis of an electronic throttle body, *IEEE/ASME Transactions on Mechatronics*, 19, 524–537, 2014.
- [15] Natori, K.; Oboe, R.; Ohnishi, K. (2008); Stability analysis and practical design procedure of time delayed control systems with communication disturbance observer, *IEEE Transactions on Industrial Informatics*, 4, 185–197, 2008.
- [16] Natori, K.; Ohnishi, K. (2008); A design method of communication disturbance observer for time-delay compensation, taking the dynamic property of network disturbance into account, *IEEE Transactions on Industrial Electronics*, 55, 2152–2168, 2008.
- [17] Navet, N.; Song, Y.; Simonot-Lion, F.; Wilwert, C. (2005); Trends in automotive communication systems, *Proceedings of the IEEE*, 93, 1204–1223, 2005.
- [18] Normey-Rico, J.E.; Camacho, E.F. (2007); Control of dead-time processes, *Advanced Textbooks in Control and Signal Processing*, Springer-Verlag London, 2007.
- [19] de Oliveira Souza, F.; Mozelli, L.A.; de Oliveira, M.C.; Palhares, R.M. (2016); LMI *International Journal of Control*, 89(10), 1962–1971, 2016.
- [20] Pan, Y.; Ozguner, U.; Dagci, O.H. (2008); Variable-structure control of electronic throttle valve, *IEEE Transactions on Industrial Electronics*, 55, 3899–3907, 2008.
- [21] Panzani, G.; Corno, M.; Savaresi, S.M. (2013); On adaptive electronic throttle control for sport motorcycles, *Control Engineering Practice*, 21(1), 42 – 53, 2013.
- [22] Pavković, D.; Deur, J.; Janszb, M.; Perić, N. (2006); Adaptive control of *Control Engineering Practice*, 14(2), 121 – 136, 2006.
- [23] Podivilova, E.; Vargas, A.N.; Shiryaev, V.; Acho, L. (2016); Set-valued estimation of switching linear system: an application to an automotive throttle valve, *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 29(4), 755–762, 2016.
- [24] Pujol, G.; Vidal, Y.; Acho, L.; Vargas, A.N. (2016); Asymmetric modelling and control of an electronic throttle, *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 29, 192–204, 2016.

- [25] Repele, L.; Muradore, R.; Quaglia, D.; Fiorini, P. (2014); Improving performance of networked control systems by using adaptive buffering, *IEEE Transactions on Industrial Electronics*, 61, 4847–4856, 2014.
- [26] Rossi, C.; Tilli, A.; Tonielli, A. (2000); Robust control of a throttle body for drive by wire operation of automotive engines, *IEEE Trans. Control Syst. Technol.*, 8(6), 993–1002, 2000.
- [27] Sarkar, B.; Chakrabarti, A.; Ananthasuresh, G.K. (2017); Synthesis of feedback-based design concepts for sensors, *Research in Engineering Design*, 131–151, 2017.
- [28] Vargas, A.N.; Menegaz, H.M.T.; Ishihara, J.Y.; Acho, L.: (2016); Unscented Kalman filters for estimating the position of an automotive electronic throttle valve, *IEEE Transactions on Vehicular Technology*, 65(6), 4627–4632, 2016.
- [29] Vasak, M.; Baotic, M.; Petrovic, I.; Peric, N. (2007); Hybrid theory-based time-optimal control of an electronic throttle, *IEEE Transactions on Industrial Electronics*, 54, 1483–1494, 2007.
- [30] Yuan, X.; Wang, Y. (2009); A novel electronic-throttle-valve controller based on approximate model method, *IEEE Trans. Industrial Electronics*, 56(3), 883–890, 2009.
- [31] Zhang, S.; Yang, J.J.; Zhu, G.G. (2015); LPV modeling and mixed constrained  $H_2/H_\infty$  control of an electronic throttle, *IEEE/ASME Transactions on Mechatronics*, 20, 2120–2132, 2015.

# Moving Object Detection and Tracking using Genetic Algorithm Enabled Extreme Learning Machine

G. Jemilda, S. Baulkani

## G. Jemilda\*

Jayaraj Annapackiam CSI College of Engineering, Nazareth, India.

\*Corresponding author: jemildajeba@yahoo.com

## S. Baulkani

Government College of Engineering, Tirunelveli, India.

ramabaulkani@yahoo.co.in

**Abstract:** In this proposed work, the moving object is localized using curvelet transform, soft thresholding and frame differencing. The feature extraction techniques are applied on to the localized object and the texture, color and shape information of objects are considered. To extract the shape information, Speeded Up Robust Features (SURF) is used. To extract the texture features, the Enhanced Local Vector Pattern (ELVP) and to extract color features, Histogram of Gradient (HOG) are used and then reduced feature set obtained using genetic algorithm are fused to form a single feature vector and given into the Extreme Learning Machine (ELM) to classify the objects. The performance of the proposed work is compared with Naive Bayes, Support Vector Machine, Feed Forward Neural Network and Probabilistic Neural Network and inferred that the proposed method performs better.

**Keywords:** curvelet transform, speeded up robust features, enhanced local vector pattern, histogram of gradient, extreme learning machine, genetic algorithm.

## 1 Introduction

Tracking an object (or multiple objects) on an image plays a vital role in the field of computer vision. Numerous applications of object tracking in computer vision are video compression, video surveillance, visual interface, man-and-computer management, medicine, augmented reality and robotics. Quick movements of objects change shapes, scenes, unstructured structures of objects and cameras have trouble tracking objects. Tracking is normally performed in the context of a program that requires the location and/or shape of the object in each frame. One of the most advanced algorithms for feature selection is the genetic algorithm. This is a stochastic method for function optimization based on the mechanics of natural genetics and biological evolution.

Faheem and Gungor [5] proposed a novel dynamic clustering based energy efficient and QoS aware routing protocol (called EGRP) which is made by the real behavior of the bird mating optimization (BMO) for smart grid applications. Fadel et al. [3] introduced Cardio radio sensor to serve as a reliable, robust and efficient communication address. In this paper, honey bee mating and cooperative channel assignment algorithms have been proposed. This significantly decreases the probability of packet loss, energy consumption, improving the life time of CRSNs in smart grids. Faheem et al. [4] discussed a novel nature-inspired evolutionary link quality-aware queue-based spectral clustering routing protocol for UASN-based underwater applications. To overcome the challenges faced by the commonly used UWSN performance indicator and to overcome the inefficiencies existing clustering based routing protocol, a novel QoS aware evolutionary cluster based protocol mentioned by Faheem et al. [6]. The main advantage is that it is robust and efficient. It is used for optimization, searching, feature extraction, segmentation and classification. This genetic algorithm reduces the processing time and increases the accuracy.

The main contributions of this paper can be described as follows:

- The use of color information (instead of grey levels only), during the entire detection process, which improves sensitivity;
- The use of background subtraction function is a good approximation of a mode function and assures high responsiveness to background changes and good accuracy;
- The inclusion of knowledge-based feedback from the soft thresholding segmentation module abates false positives and prevents from deadlock situations in background update;
- The system is conceived to be auto-adaptive, i.e., to work with the best possible performances on all possible scenarios. Results show that accurate results are aligned with other state-of-the-art approaches, but they are more stable in all corner-case situations and higher frame-per-second rates.
- It also shows very limited memory (of all types) requirements, as it is usually demanded for embedded applications.

The rest of this paper is organized as follows: Section 2 discusses the related work. Section 3 explains the proposed method including its design idea and practical implementation approach. Section 4 presents simulation parameters. Section 5 provides experimental results where the effectiveness of the proposed work is compared to the existing methods. Finally, Section 6 concludes this paper by summarizing our results, significance and future possibilities of the work.

## 2 Related work

A new tool developed by Arnab Roy et al. [12] to speed up the dramatic change of detection of low-end laptop webcam stuff with specific hardware for processing high-speed images. The first algorithm is fast and ensures the second algorithm edge of the object found more clearly at the expense of slow image processing. This method is used to reduce the average delay in motion of the objects up to 45.5% and the memory usage by about 14%, while keeping the same accuracy. Hsu-Jung Cheng et al. [2] proposed a framework for detecting vehicles in aeronautical surveillance using the Dynamic Bayesian Network and found that this method has flexibility and good generalization capabilities. Felix M. Philip et al. [11] proposed visibility model based on the second derivative to predict the objects, spatial tracking model and tangential weighted function to track multiple objects. But tracking multiple objects in low resolution videos is not possible.

## 3 Methodology

This section briefly explains about the methods used in this paper for tracking the objects and it was shown in the Figure 1.

### 3.1 Input video

The input video is chosen as it consists of the objects to track. The objects in the video must be moving, since this project is to capture the moving objects in the video.

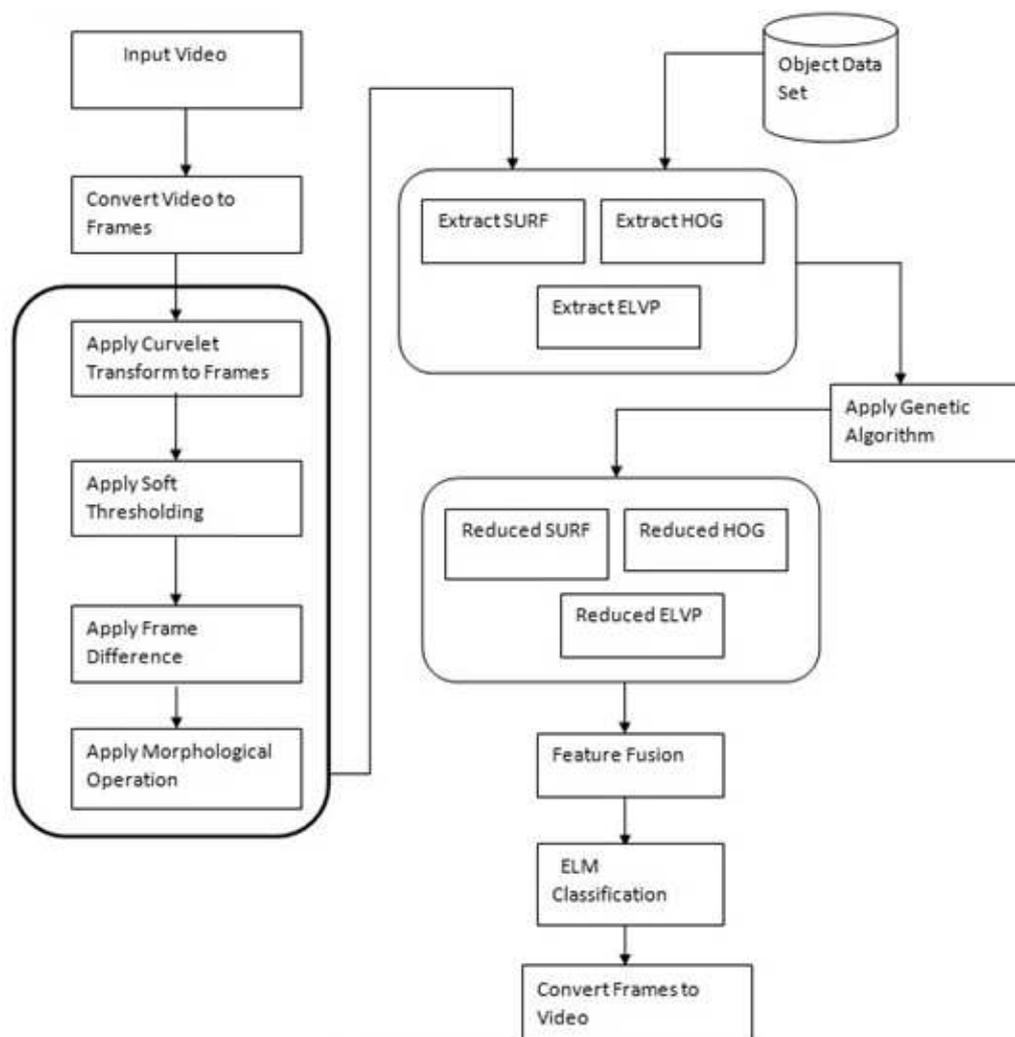


Figure 1: Architecture diagram

### 3.2 Divide video into image frames

Video technology is widely used for electronically capturing, recording, processing, storing, transmitting and analyzing a sequence of still scenes showing images in motion. The number of still pictures of videos per unit of time ranges from six or eight frames per second (frame/s) to 120 frames per second or more. These frames are saved and processed consequently. In this context, the video is divided into number of frames for example, the video of up to 10 seconds is break down into 80 to 85 frames.

### 3.3 Object localization

The moving object is localized from each and every frame by using background subtraction algorithm. The traditional background subtraction algorithm is modified and a new concept is implemented. The curvelet transform is applied on each and every frame. Then soft thresholding is applied to remove noise. Next, frame difference process is applied to segment the moving object and the background. Finally, mathematical morphology concept is applied to fill the gap and edges in the detected objects.

### **Curvelet transform**

Curvelet is not a coordinate technique for presenting multiple objects. It differs from the other wave in the direction of transition to the extent of oriented translation. The use of this transform is optimally sparse representation of objects with edges and optimal image reconstruction in ill-posed problems [9].

### **Soft thresholding**

Soft thresholding is a popular tool in computer vision and machine learning. It is used to remove noise from an image [1].

### **Frame difference approach**

Frame differencing is a technique used to find the difference between two video frames. It is used to segment the moving object from the background. If the pixels have changed, apparently there was something changed in the image. The algorithm of frame difference is relatively simple. It is described in very detailed manner as follows:

1. Convert the incoming frame to grayscale.
2. Subtract the current frame from the background model, which is the previous frame.
3. For each pixel, if the difference between the current frame and background is greater than a threshold then the pixel is considered as part of the foreground. Otherwise it is considered as the background.
4. The foreground pixel is denoted as white color and the background pixel is denoted as the black color. Finally the output image is produced.

### **Mathematical morphology**

Mathematical morphology is a powerful tool for extracting structural characteristics in an image and is useful for characterizing shape information. The obtained segmented object may include a number of disconnected edges due to non-ideal segmentation of moving object edges. Therefore, some morphological operation is needed to generate connected edges. Here, a binary closing morphological operation is used and it is determined by a structuring element. The effect of the operator is to preserve background regions that have a similar shape to this structuring element, or that can completely contain the structuring element, while eliminating all other regions of background pixels [8].

## **3.4 Object identification**

After locating the moving object, the next step is to identify the object. To do this first the feature extraction techniques are applied on to the localized object. In this paper the texture, color and shape information of objects are considered.

### **Shape feature extraction using SURF**

Speeded Up Robust Features (SURF) is a local feature detector and descriptor that can be used for tasks such as object recognition, registration, classification and 3D reconstruction. It is partly inspired by the scale-invariant feature transform (SIFT) descriptor. The standard version of SURF is several times faster than SIFT and claimed by its authors to be more robust against different image transformations than SIFT. SURF descriptors can be used to locate and

recognize objects, people or faces, to make 3D scenes, to track objects and to extract points of interest [15].

SURF uses the determinant of Hessian for selecting the scale, as it is done by Lindeberg. Given a point  $p=(x, y)$  in an image  $I$ , the Hessian matrix  $H(p, \sigma)$  at point  $p$  and scale  $\sigma$ , is defined as follows:

$$H(p, \sigma) = \begin{pmatrix} L_{xx}(p, \sigma) & L_{xy}(p, \sigma) \\ L_{xy}(p, \sigma) & L_{yy}(p, \sigma) \end{pmatrix} \quad (1)$$

where  $L_{xx}(p, \sigma)$  etc. are the second-order derivatives of the grayscale image. The box filter of size  $9 \times 9$  is an approximation of a Gaussian with  $\sigma=1.2$  and represents the lowest level (highest spatial resolution) for blob-response maps. The pseudo code for SURF is shown in Algorithm 1.

Algorithm 1: Pseudo Code of SURF algorithm
Input : Input Image Output : SURF Key Points
<ol style="list-style-type: none"> <li>1. Begin</li> <li>2. For all pixel <math>i=1 : P</math></li> <li>3. do</li> <li>4. Compute the gradient.</li> <li>5. Compute gradient magnitude and orientation</li> <li>6. Find the scalespace response using DoH filters with different <math>\sigma</math>.</li> <li>7. Find local maxima LM with different scales and octaves.</li> <li>8. if LM==Highest Maxima value</li> <li>9. Select as SURF Key Points</li> <li>10. end if</li> <li>11. end</li> <li>12. End</li> </ol>

### Colour feature extraction using HOG

The histogram of oriented gradients (HOG) is a feature descriptor used in computer vision and image processing for the purpose of object detection. It is used to count the occurrences of gradient orientation in localized portions of an image [10]. The pseudo code for HOG is shown in Algorithm 2.

Algorithm 2: Pseudo Code of HOG algorithm
Input : Input Image Output : HOG Feature
<ol style="list-style-type: none"> <li>1. Begin</li> <li>2. For all block <math>i=1 : B</math></li> <li>3. do</li> <li>4. Compute gradient values in horizontal directions.</li> <li>5. Compute gradient values in vertical directions.</li> <li>6. Create histograms of gradient image</li> <li>7. end</li> <li>8. Concatenate normalized histograms from all of the block regions.</li> <li>9. End</li> </ol>

### Texture feature extraction using ELVP

Enhanced Local Vector Pattern (ELVP) is a novel vector representation developed to represent the 1D direction and structure information of local texture and the adjacent pixels with

diverse distances from different directions. Based on the vector representation, the LVP descriptor is proposed to provide various 2D spatial structures of micro patterns with various pair wise directions of vector of the referenced pixel and its neighborhoods.

The proposed local pattern descriptor encodes the LVPs by using the four pairwise directions of vector of the referenced pixel and its neighborhoods. Especially, each pairwise direction of vector of the referenced pixel generates the transform ratio which is used to design the weight vector of dynamic linear decision function for encoding the distinct 8-bit binary pattern of each LVP. Since the binary code can be considered as a two-class case by using dynamic linear decision function to calculate the Comparative Space Transform (CST) values of the neighborhoods for encoding a bit string via the sign function [7]. The pseudo code for ELVP is shown in Algorithm 3.

Algorithm 3: Pseudo Code of ELVP algorithm
Input : Input Image
Output : ELVP Feature
<ol style="list-style-type: none"> <li>1. Begin</li> <li>2. For all pixel <math>i=1 : P</math></li> <li>3. do</li> <li>4. Take the one pixel <math>G_c</math></li> <li>5. Take the pixel <math>N</math> in direction of angle <math>\beta</math></li> <li>6. Calculate the Difference <math>D</math> between <math>G_c</math> and <math>N</math>  <math display="block">V_{\beta,D}(G_c) = (I(G_{\beta,D}) - I(G_c))</math> </li> <li>7. Calculate the binary pattern  <math display="block">LVP_{p,R}(G_c) = \{LVP_{p,R,\beta}(G_c)   \beta = 0^\circ, 45^\circ, 90^\circ, 135^\circ\}</math> </li> <li>8. Calculate the binary description and store it into the bin value.</li> <li>9. Store Description in the bin array</li> <li>10. End</li> <li>11. End</li> </ol>

### 3.5 Feature selection

The genetic algorithm (GA) is used for finding the best features from the SURF, HOG and ELVP. A better solution can be evolved using a population of strings (called chromosomes or the genotype of the genome) and by encoding candidate solutions (called phenotypes) to an optimization problem. Solutions are represented in binary as strings of 0s and 1s. The evolution usually starts from a population of randomly generated individuals and happens in generations. In each generation, the fitness of every individual in the population is evaluated. Multiple individuals are selected from the current population (based on their fitness), and modified (recombined and possibly randomly mutated) to form a new population. The new population is then used in the next iteration of the algorithm. The algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population. If the algorithm has terminated due to a maximum number of generations, a satisfactory solution may or may not have been reached. GA proceeds to initialize a population of solutions randomly, and then improve it through repetitive application of mutation, crossover, and inversion and selection operators [13].

Apply the genetic algorithm for each feature extraction method such as SURF, HOG and ELVP. After applying the genetic algorithm the reduced SURF, reduced HOG and reduced ELVP is produced.

### 3.6 Feature fusion

The next step is to combine the reduced SURF, reduced HOG and reduced ELVP into single feature vector. To do this, concatenate all features into a single array. Finally the single feature vector is produced.

### 3.7 Object classification

A new learning algorithm called extreme learning machine (ELM) is used to classify objects for single-hidden layer feed forward neural networks (SLFNs) which randomly chooses hidden nodes. This algorithm provides good performance at fast learning speed. The ELM is an emerging learning technique provides efficient solutions to generalized feed-forward networks including (both single and multi-hidden-layer) neural networks, radial basis function (RBF) networks, and kernel learning. ELMs have classification capability. ELMs have significant features like fast learning speed, ease of implementation, and minimal human intervention. They thus have strong potential as a viable alternative technique for large-scale computing and machine learning [14].

### 3.8 Generating video from frames

To generate a video from a set of sequences or set of frames, it starts with the number zero. This will work as long as the sequence is unbroken. The frame rate of the resulting video is 3 frames per second so that each still can be seen for a short period of time. The rescale of the picture to the desired resolution can be done to manage the size of the resulting video.

### 3.9 Output video

The output video is generated from the filtered and classified frames. These frames are combined to form the output video.

## 4 Simulation

The Simulation results on five different sequences recorded with three different platforms is presented in Table 1. In all cases, the sensor was a pair of forward-looking AVT Marlin F033C cameras, which deliver synchronised video streams of resolution 640 x 480 pixels at 13-14 frames per second. Bahnhofstrasse (999 frames) and Linthescher (1208 frames) have been recorded with a child stroller (baseline  $\approx 0.4$  m, sensor height  $\approx 1$  m, aperture angle  $\approx 65^\circ$ ) in busy pedestrian zones, with people and street furniture frequently obstructing large portions of the field of view. Loewenplatz (800 frames), Bellevue (1500 frames) and City (3000 frames) have been recorded from a car (baseline  $\approx 1$  m, sensor height  $\approx 1.3$  m, aperture angle  $\approx 50^\circ$ ) driving on inner-city streets among other cars, trucks and trams. Pedestrians appear mostly on sidewalks and crossings, and are observed only for short time spans. Lighting and contrast are realistic, with most sequences recorded on cloudy days.

## 5 Experimental images

The datasets used in this work are taken from <https://motchallenge.net/vis/MOT17-13-SDP> and <https://motchallenge.net/vis/ETH-Crossing>. Only the frames of the datasets for various movements of the moving object 1 and 2 are shown in the Figure 2 and Figure 3.

Table 1: Simulation results

Simulation Parameter	Value
Population	12
No of Generation	10
Video Resolution	640 x 480
Background updation Frame	10



Figure 2: Various movements of the moving object 1

### 5.1 Experimental results

To evaluate the performance of the proposed system, this paper used the dataset with various objects like moving objects, lemming, occlusion, multiple objects, and indoor objects. To examine the effectiveness of the proposed system, it is compared with various performance metrics. The various performance metrics used to compare the effectiveness are average center error, overlap rate, Detection Accuracy, Sensitivity, Specificity, F-Measure, Detection Error, Execution Time, Precision Rate and Recall Rate.

#### Overlap rate

It shows the overlapping of each frame by frame.

$$OverlapRate = \frac{area (R_T \cap R_G)}{area (R_T \cup R_G)} \quad (2)$$

Where  $R_T$  - Tracking result of each frame  
 $R_G$  - Corresponding ground truth

#### Average center error rate

It shows the error rate for each frame

$$AverageCenterErrorRate = area (R_T \cup R_G) - area (R_T \cap R_G) \quad (3)$$

Where  $R_T$  - Tracking result of each frame  
 $R_G$  - Corresponding ground truth



Figure 3: Various movements of the moving object 2

### Precision rate

The precision is the fraction of retrieved instances that are relevant to the find.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

Where TP - True Positive (Equivalent with Hit)

FP - False Positive (Equivalent with False Alarm)

### Detection accuracy

It is one of the parameter which is used to analyse the performance of the proposed method.

$$\text{Detection Accuracy} = \frac{\text{No. of Correctly Classified Objects}}{\text{Total No. of Objects}} \quad (5)$$

### F-measure

F-measure is the ratio of product of precision and recall to the sum of precision and recall. The f-measure can be calculated as

$$F_m = (1+a) * \frac{\text{Precision} * \text{Recall}}{a * \text{Precision} + \text{Recall}} \quad (6)$$

Where  $a$  - real value

### Sensitivity

Sensitivity also called the true positive rate or the recall rate measures the proportion of actual positives. It is calculated using

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (7)$$

Where TP - True Positive (equivalent with hit)

FN - False Negative (equivalent with miss)

### Specificity

Specificity measures the proportion of negatives which are correctly identified. It is calculated using

$$\text{Specificity} = \frac{\text{TN}}{(\text{FP} + \text{TN})} \quad (8)$$

Where TN - True Negative (equivalent with correct rejection)

FP - False Positive (equivalent with false alarm)

### Detection error

It is one of the parameter which is used to analyse the performance of the proposed method. It is calculated using the below said formula.

$$\text{Detection Error} = \frac{\text{No. of Wrongly Classified Objects}}{\text{Total No. of Objects}} \quad (9)$$

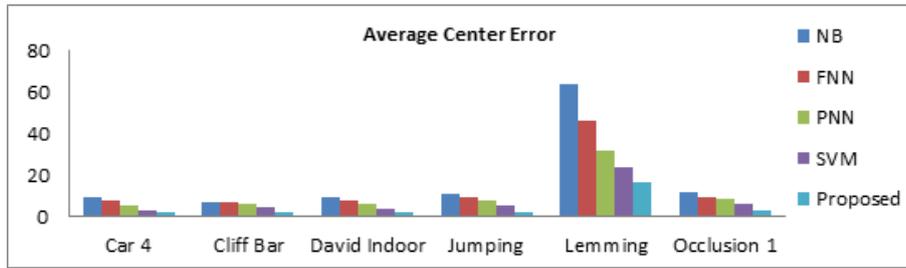


Figure 4: Average center error for various methods with different datasets

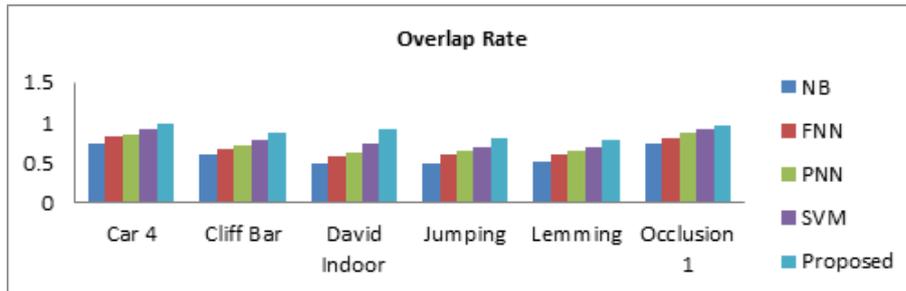


Figure 5: Overlap rate for various methods with different datasets

### Execution time

It is one of the parameter which is used to analyse the performance of the proposed method. It is calculated using the below said formula.

$$\text{Execution Time} = \text{Ending Time} - \text{Starting Time} \quad (10)$$

To analyse the performance of the proposed system, it is compared with the above mentioned performance metrics. The performance comparison of the average center error value of the proposed method and other four existing approaches such as NB, FNN, PNN and SVM are plotted in the graphs as below.

In the Figure 4, the average center error value of the five methods including the proposed method is compared. The average center error value of the proposed method is lower than the other four existing approaches. Because of the low average center error value, the proposed method is better than the other four existing approaches.

The performance comparison of the overlap rate of the proposed method and other four existing approaches such as NB, FNN, PNN and SVM is compared and shown in the Figure 5. The overlap rate value of the proposed method is higher than the other methods. Because of the high overlap rate value, the proposed method is better than the other four existing approaches.

The performance precision rate of the proposed method and other four existing approaches such as NB, FNN, PNN and SVM is plotted in the Figure 6. The precision rate value of the proposed method is higher than the other four existing approaches. Because of the high precision rate value, the proposed method is better than the other four existing approaches.

The detection accuracy value of the proposed method is higher than the other four existing approaches as shown in the above Figure 7. Because of the higher detection accuracy value, the proposed method is better than the other four existing approaches.

The sensitivity value of the proposed method is higher than the other four existing approaches as shown in the above Figure 8. Because of the high sensitivity value, the proposed method is better than the other four existing approaches.

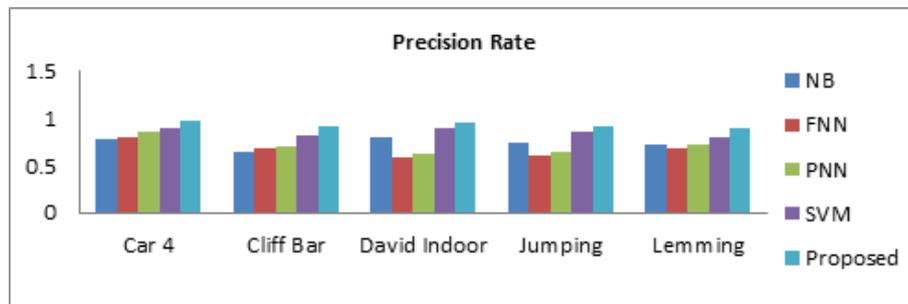


Figure 6: Precision rate for various methods with different datasets

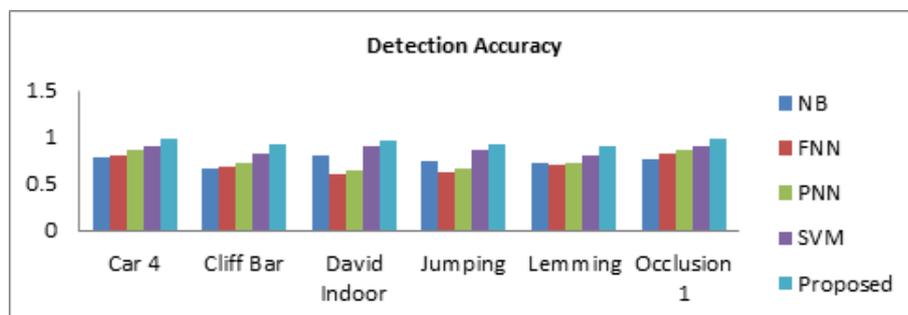


Figure 7: Detection accuracy for various methods with different datasets

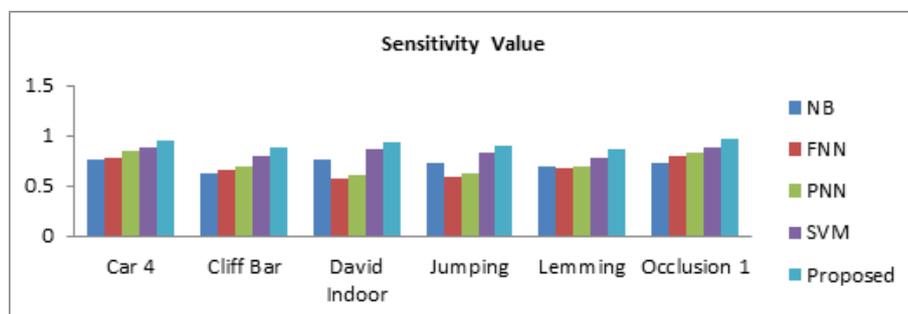


Figure 8: Sensitivity value for various methods with different datasets

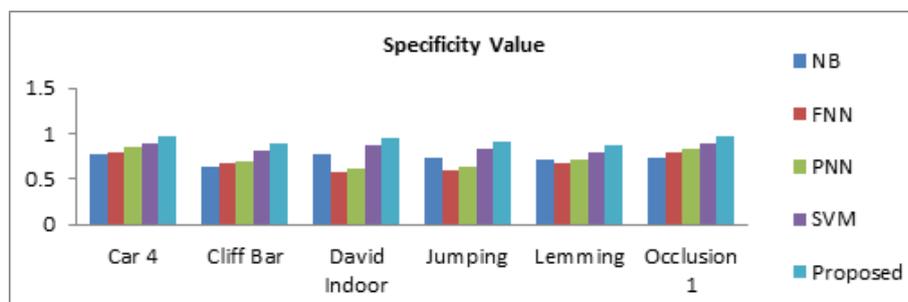


Figure 9: Sensitivity value for various methods with different datasets

The Figure 9 shows the specificity value of the proposed method is higher than the other four existing approaches. Because of the high specificity value, the proposed method is better than the other four existing approaches.

## 6 Conclusion

In this paper, the moving objects in the video are tracked and it is highlighted to show that the object is moving in the video. For the first the video is divided into frames. First the input given is a video, which is divided into frames. Then the moving object is localized from each and every frame by using background subtraction algorithm. After locating the moving object the next step is to identify the object. To do this first the feature extraction techniques are applied on to the localized object. After extracting all the features these feature are reduced using the genetic algorithm and then these reduced feature set are fused to form a single feature vector. Finally these feature vectors are given into the Extreme Learning Machine (ELM) to classify the objects. The classified frames are then combined to form a video. In these output video, the moving object is highlighted. To know the performance of the proposed method, we compared the results with the various methods. From the experiment results, our proposed methods shows better results for identifying the moving object in the video file and are very efficient and reliable.

As a future work, an algorithm that detects the tracking failure and recovers the tracking process when the target tracking fails due to long duration of heavy occlusion is planned.

## Bibliography

- [1] Biswas, M.; Om H. (2012); A new soft thresholding Image Denoising method, *Science Direct*, 6:10-15,2012.
- [2] Cheng, H.-Y.; Weng, C.-C.; Chen Y.-Y.(2012); Vehicle Detection in Aerial Surveillance Using Dynamic Bayesian Networks, *IEEE Transactions on Image Processing*, 21(4): 2152-2159, 2012.
- [3] Fadel, E.; Faheem, M.; Gungor, V.; Nassef, L.; Akkari, N.; Malik, M. (2017); Spectrum-Aware Bio-Inspired Routing in Cognitive Radio Sensor Networks for Smart Grid Applications *Computer Communications*, 106-120, 2017.
- [4] Faheem, M.; Tuna, G.; Gungor, V.C. (2016); LRP: Link quality- aware queue- based spectral clustering routing protocol for underwater acoustic sensor networks, *International Journal of Communication Systems*, 2016.
- [5] Faheem, M.; Gungor V.C. (2017); Energy Efficient and QoS-aware Routing Protocol for Wireless Sensor Network-based Smart Grid Applications in the Context of Industry 4.0, *Applied Soft Computing*, 1-13, 2017.
- [6] Faheem, M.; Tuna, G.; Gungor V.C. (2017); QERP: Quality-of-Service (QoS) Aware Evolutionary Routing Protocol for Underwater Wireless Sensor Networks, *IEEE Systems Journal*, 2017.
- [7] Fan, K.-K.; Hung, T.-Y. (2014); A Novel Local Pattern Descriptor-Local Vector Pattern in High-Order Derivative Space for Face Recognition, *IEEE Transactions on Image Processing*, 23(7), 2877 - 2891, 2014.

- [8] Kimori, Y. (2013); Morphological image processing for quantitative shape analysis of biomedical structures: effective contrast enhancement, *Journal of Synchrotron Radiation*, 1(20), 848-853, 2013.
- [9] Kourav, A.; Singh P. (2013); Review on curvelet transform and its applications, *Asian Journal of Electrical Sciences*, 2(1): 9-13, 2013.
- [10] Li, Y.; Su G. (2015); Simplified histograms of oriented gradient features extraction algorithm for the hardware implementation, *International Conference on Computers, Communications and Systems (ICCCS)*, 192 -195, 2015.
- [11] Philip, F.M.; Mukesh R.(2016); Hybrid tracking model for multiple object videos using second derivative based visibility model and tangential weighted spatial tracking model, *International Journal of Computational Intelligence Systems*, 9(5): 888-899, 2016.
- [12] Roy, A.; Shinde,S.; Kang, K.-D. (2012); An Approach for Efficient Real Time Moving Object Detection, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 5(3), 2012.
- [13] Shingade, A.; Ghotkar A.(2014); Survey of Object Tracking and Feature Extraction Using Genetic Algorithm, *International Journal of Computer Science and Technology*, 5(1), 2014.
- [14] Wang, Y.; Cao, F.; Yuan, Y. (2014); A Study on Effectiveness of Extreme Learning Machine, *arXiv:1409.3924v1 [cs.NE]*, 13, 2014.
- [15] Zohrevand, A.; Ahmadyfard, A.; Pouyan, A.; Imani, Z. (2014); A SIFT based object recognition using contextual information, *Iranian Conference on Intelligent Systems (ICIS)*, 1-4, 2014.

# A Hybrid Failure Diagnosis and Prediction using Natural Language-based Process Map and Rule-based Expert System

D. Kim, Y. Lin, S. Lee, B.H. Kang, S.C. Han

## Dohyeong Kim, Sungyoung Lee

Department of Computer Science and Engineering, Kyung Hee University,  
1732, Deogyong-daero, Giheung-gu, Yongin-si Gyeonggi-do 17104, Republic of Korea  
dhkim@oslab.khu.ac.kr, sylee@oslab.khu.ac.kr

## Yingru Lin, Byeong Ho Kang

School of Engineering and ICT,  
University of Tasmania,  
Private Bag 87, Hobart, Tas 7001, Australia  
Yingru.Lin@utas.edu.au, Byeong.Kang@utas.edu.au

## Soyeon Caren Han\*

School of Information Technologies, University of Sydney,  
Sydney, NSW, Australia

\*Corresponding author: Caren.Han@sydney.edu.au

**Abstract:** Preventive maintenance is required in large scale industries to facilitate highly efficient performance. The efficiency of production can be maximized by preventing the failure of facilities in advance. Typically, regular maintenance is conducted manually in which case, it is hard to prevent repeated failures. Also, since measures to prevent failure depend on proactive problem-solving by the facility expert, they have limitations when the expert is absent, or any error in diagnosis is made by an unskilled expert. In many cases, an alarm system is used to aid manual facility diagnosis and early detection. However, it is not efficient in practice, since it is designed to simply collect information and is activated even with small problems. In this paper, we designed and developed an automated preventive maintenance system using experts' experience in detecting failure, determining the cause, and predicting future system failure. There are two main functions in order to acquire and analyze domain expertise. First, we proposed the network-based process map that can extract the expert's knowledge of the written failure report. Secondly, we designed and implemented an incremental learning rule-based expert system with alarm data and failure case. The evaluation results shows that the combination of two main functions works better than another failure diagnosis and prediction frameworks.

**Keywords:** expert's knowledge, preventive maintenance, failure prediction, alarm management, knowledge reuse.

## 1 Introduction

When a failure or error occurs in massive industrial facilities, it could cause industrial disaster or the factory operation shutdown, which brings enormous financial and social consequences [13]. Therefore, to maintain the normal workflow and maximize the profit, companies are running safety management processes, working with best efforts for the preventive maintenance and failure prediction. Preventive maintenance is the most important process to prevent failure before it occurs, and it aims to maintain the facilities in the best condition [9].

However, despite the periodic care, preventive maintenance cannot stop the continuous failures, since there are facilities where periodic maintenance is not effective enough. Furthermore, excessive maintenance without considering the cause behind the malfunction is inefficient and economically unproductive. After a failure occurs, it is necessary to take proper actions as soon

as possible not only to restart the operation, but also to recover the rate of operation to avoid further financial damage. However, these actions are entirely dependent on the skilled and experienced human experts. The main problem here are that there is a limited number of skilled human experts, and the failure rate of such actions rises dramatically with the absence of such experts. Furthermore, the experts treat these errors based on experiences which may not always be recalled correctly. Their experiences may also not cover all sorts of problems as the system is prone to unknown failures. The possibility of human errors cannot be waived too. This makes depending on only expert experience undesirable [7].

Alarm system supports the facility maintenance by collecting the status of the facility from the installed sensors in real time and providing collected information to the manager in the control room. However, continuous operation of the facility causes deterioration, exchange of the components, and requires additional sensors. It changes the design of the initial alarm system and brings new types of alarms and occurrence patterns. Such phenomenon is known as alarm flooding which makes field workers ignore the alarms as the excessive amount of information cannot be handled. Many research projects have been undertaken to solve this problem, but most of them are about solving the alarm flooding problem. They focused on reducing the number of alarms to an affordable level, by extracting unusual alarms with a categorization of alarm types and analysis of the alarm patterns using statistical models. The limitation of such projects is that even if they succeed and the alarms are provided in manageable level, there are many additional works which require enormous time and efforts such as analysis of alarms, diagnosis, and failure prediction [1].

Since the cause of those system failures are not one dimensional, knowledge and experiences of an individual human expert is not enough to generalize the failure knowledge. Hence, failure knowledge must be built based on the continuously accumulated knowledge from several experts. However, this takes a lot of time that results in actions being delayed. Even when we can identify the abnormal facility status, financial damages are inevitable. In this paper, we propose a preventive maintenance system which finds the abnormal status in massive industrial facilities such as steelworks, and automatically provides the knowledge required for the maintenance and prevention of failures. We developed an expert system that recognizes the status of the facilities by using alarm data which can be used to find the abnormal status of the facilities. In this system, expertise regarding the alarms is arranged in a systematic structure, so that the system could recognize the status of the facilities with automatically generated alarms.

Furthermore, we also developed a system which contributes to the knowledge base by analysing failure reports so that it can reuse the accumulated failure records. Failure to reports include a description of analysis, diagnosis and actions taken related to the failure occurred. Thus, it contains knowledge regarding the causal relationship of failures as well as the implicit knowledge of experts regarding every step of the actions taken. This makes the failure report a primary source of solving the problem when failure occurs. Even though the failure circumstance is documented as a report, the report is still not systematically managed and well obtained the the relevant information.

A proper understanding the casual relationship of failure circumstances can enable us to diagnose the failure and to predict when the same problem occurs. We analyzed and processed failure reports which are composed of natural language, and converted them into structured knowledge, called as "knowledge map", for systematic storage and management. The knowledge about the alarm and failure reports are designed to have consistency in format so that they can operate interactively. When an alarm occurs, the status of alarm is detected and mapped with the process map by the expert system, which enables us to recognize the problem based on the status of facilities and to diagnose the process which made the facility status abnormal.

## 2 Related works

There are numerous researches related to alarm, thanks to its usefulness in maintaining and predicting faults of facilities in various fields. Outlier detection methods which are based on the analysis of alarm data aim to find the abnormal patterns with outlier values [7, 12].

However, such methods are not available in real-time, since they require to analyze massive amount of data, and interception of experts is also required in order to determine that the found abnormal alarms are indeed useful information related to the malfunctioning. There are also researches based on the machine learning, such as predicting alarms and temperature level of the steel production facilities using LS-SVM [8].

Knowledge based ALAP system is a rule based expert system, which aims to diagnose the faults with alarms [4]. Furthermore, a research has been done, which tries to handle the alarms efficiently by prioritizing the meaningful alarms from the alarm flooding using fuzzy logic [15]. There are also researches which focus on the design of alarm system.

In the research of [3], an evaluation of the alarm filters which are suggested in [1] to find the optimal ones. In the research of [1] clustering of alarm is done by evaluating the consecutive occurrence pattern of alarms. Also, various alarm management techniques including frameworks, data filtering, alarm delay and alarm deadlines are suggested, in order to design an optimal alarm system [5]. The researches which converge different methodologies include [16], which aims to manage alarm flooding by using both dynamic alarm management and Bayesian measurement. Such researchers are focusing on the alarm management and features analysis, and are limited in a sense that they are not considering the actual relation between the faults and alarms, and whether the managers are actually interested in the alarm.

In the field of natural language processing, the researches which extract knowledge from document with natural language have been done actively. Researches for extracting information, including specific terminology, objects, and concepts using machine learning, rule base, dictionary and corpus are, having been performed continuously, and among them there are researchers which especially interested in extracting entities, a relationship between entities, and occurring events from the medical documents and newspapers [6, 10, 11, 14]. However, such researches are restricted to simply extract the information from the unstructured knowledge written in natural language. However, proposing fault analysis system is capable of reconstructing the extracted information into structured knowledge, which enables the systematic management of knowledge.

Previous researches for alarm processing mainly try to recommend meaningful alarm among the massive amount of alarms, or reduce the number of alarms to make it easy to take action after monitoring. Such researches suffer from high dependency on human experts, since the experts have to analyze the meaning of alarms after the alarm processing step after all. Proposing a system constructs an expert system on the relationship between alarms and facility, which enables to automatically recognize the status of a facility based on the occurring alarms. Furthermore, we converted the historic records of fault reports into structured system knowledge which include causal relationships. The structure of such knowledge is designed to be interoperable with the knowledge related to alarm. Therefore, fault diagnosis can be done by inferencing the status of the facility, mapping the inferred status to previous fault records, and tracking the problem occurred in the mapped case using causal relationship knowledge.

Prediction of fault which will happen after the current alarm can be done also, by reviewing the problem which occurred after the mapped previous case. Namely, the series of processes including diagnosis of the facility based on alarms, and fault diagnosis/prediction with the previous fault history is done automatically by the system.

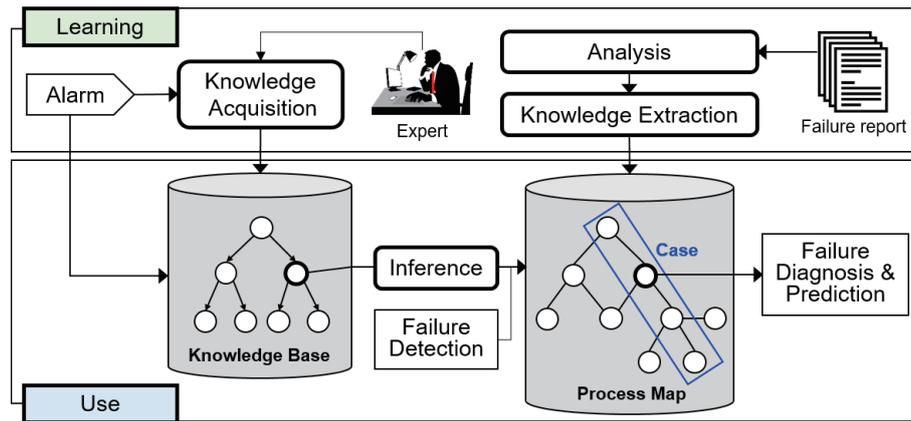


Figure 1: Overview of the proposed preventive maintenance system

### 3 Proposed preventive maintenance system

In this section, the architecture of proposing system and the main components are described. The proposed system is composed of two closely interoperating main components, a rule-based expert system for processing alarms and a failure analysis system to predict future system failures. A knowledge base for finding the abnormal status of the facility and preventive or diagnosis knowledge base about abnormal signs are essential to perform preventive maintenance automatically with the system. Those two kinds of knowledge come from distinct sources, but they share equivalent unit for the failure diagnosis and prediction. The expert system generates and manages the knowledge for finding the failure signs. The failure analysis system extracts knowledge from the failure reports, converts into causal relationship knowledge for failure diagnosis and prediction, and stores it into knowledge storage named as process map.

Figure 1 represents the preventive maintenance process via the expert system where failure prediction is done by failure prediction system. The proposed system uses alarms, which include signs of failures, and failure knowledge which includes causal relationship. All the alarms occur in real-time. Alarm system collects those alarms periodically for every hour and sends them to the expert system. An alarm, includes identification of facility in which the failure occurred, contents of the alarm, and quantitative numbers measured during an hour. Furthermore, it is stored in alarm storage for a certain period, and is used in alarm analysis and construction of the knowledge base by the field experts. The expert system possesses an inference engine which acquires the knowledge required for the recognition of problems from alarms, and knowledge acquisition engine which is used when the experts directly generate and manage knowledge in the knowledge base. Failure prediction is done using the inference engine and failure knowledge of the expert system after the alarm occurs.

The inference engine indicates towards facilities that are problematic during the occurrence of an alarm. It also searches for knowledge related to the problem, acquires and provides the cause of the problem and predicts possible additional failures. Thus, the time required for the field experts to recognize and solve the problem is greatly reduced, since the problem description, cause of the problem and viable actions are automatically provided by the system. When a failure occurs despite the preventive maintenance, the problem is diagnosed and solved by directly investigating in the field. A report of the entire process is written by the field experts.

The failure analysis system analyzes the failure reports in order to reuse expert experience knowledge. Failure reports include information related to the problem such as current status, cause, and actions taken, so that it can represent a causal relationship of the problem. Thus,

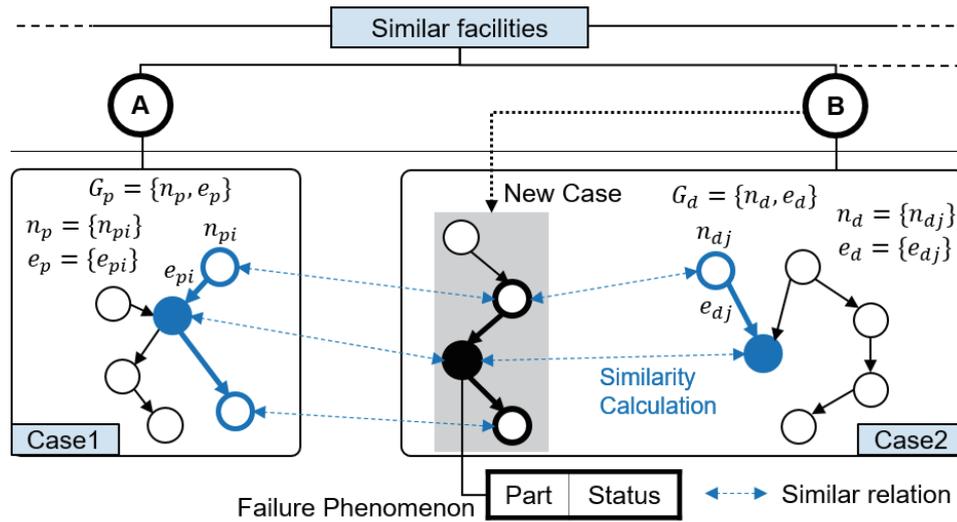


Figure 2: Conceptual design of process map

it can be used as a knowledge base for failure diagnosis and prediction. Failure reports include domain dependent terminologies, implicit representation, and are written in a unique way for each writer. This is the reason why the system utilizes not only the failure reports, but also the domain terminologies and domain knowledge. Constructed process map interoperates with the expert system to predict failure and to analyse of the cause of the failures. Figure 1 represents the conceptual process of the preventive maintenance. It represents the interaction between the expert system for alarm processing and the process map based on the failure report analysis. The failure cases which were acquired via such interaction can be used for failure diagnosis and failure prediction.

Knowledge is obtained by analyzing failure reports, extracting minimum semantic units from failure reports written in natural language, constructing causal relationships between them, and mapping the failure of the target facility. This knowledge is then labelled as failure knowledge. After that, failure knowledge is stored and managed in a process map. In the failure report, the simplest and basic type of sentence is the minimum semantic unit. Two components of the facility corresponding to the subject are extracted from the sentence: the subject that refers to the target of the facility and the predicate that means the status/operation of facility. The acquired unit knowledge with the form of node is called ‘failure phenomenon’. After extraction process, subject indicate as Part of the component and predicate indicates as the status of the component. We analyzed failure reports and extracted short sentences from unstructured natural language. Order between two short sentences and consistency of meaning is considered while constructing a relationship which is referred to as failure case. A series of processes including how the specific facility has caused the problem and how the problem is solved, is configured in the order of occurrence through the relation between the failure phenomena.

Figure 2 shows the concept of a process map. The process map is based on the facility. There are numerous facilities in the factory, and many facilities have similar functions, but the type of facilities usually varies. Therefore, managing individual knowledge of these facilities can make knowledge generation very complex and knowledge management very inefficient. The proposed process map designates representative facilities with functions and designates the facilities having similar functions after the representative facilities. Having a similar function means that the structure of facilities is similar, and the usage of facilities and the relation with other facilities are also similar. This knowledge model is suitable for complex domains such as large factories

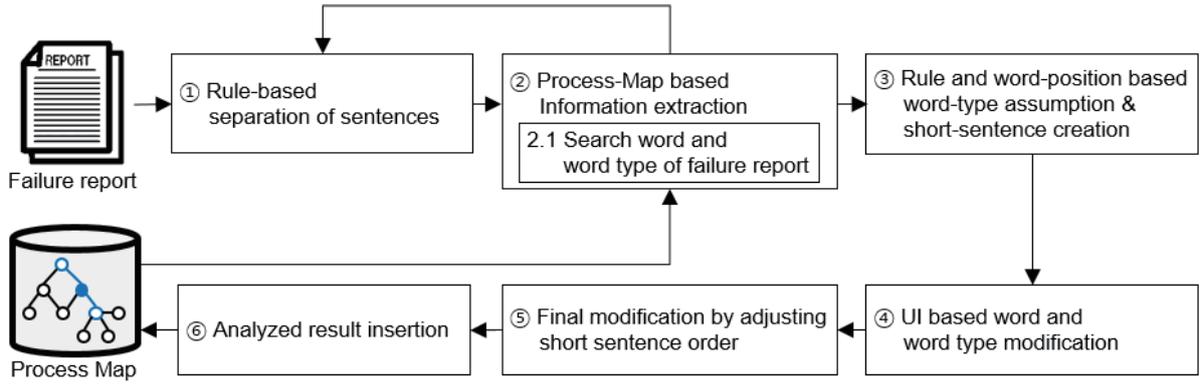


Figure 3: Process of failure report analysis

with various facilities. The failure report describes the failure analysis and failure action for the facilities that have a failure so that the failure cases analyzed from failure reports can be linked to these facilities. In Figure 2, A and B are facilities with similar functions where each facility has one or more failure cases and is represented by a single pass. If failure cases have the same failure phenomenon, failure phenomenon is expressed as multipath constructed by integrating and sharing failure phenomenon. For the similarity calculation, we applied following similarity formulation and text-based similarity algorithm:

$$Sim(n_{dj}, n_{pi}) = Sim_{Text}(n_{dj}, n_{pi}) + Sim_{Rel}(n_{dj}, n_{pi}) \quad (1)$$

$$Sim(n_{dj}, n_{pi}) = EditDistance(n_{dr}, n_{pir}) + EditDistnace(n_{djE}, n_{piE}) \quad (2)$$

Because similar failure phenomena of similar facilities have a relationship, whole structure of the process map is in a network. So, similar failure cases can be referenced in addition to direct failure case. In case that contents of failure knowledge are weak or non-existent, the process map that uses similar facilities and failure cases can easily be referenced in the indirect method. Thus, we can achieve high knowledge usability.

## 4 Process map construction

Failure reports written in natural language are transformed into structured forms using natural language processing techniques and then stored in the process map. The analysis is conducted automatically by the system and the results of that can be modified directly by the user through the user interface. The contents stored in the process map are used as information for analyzing the failure report, and then those analysed contents are connected to each other with the same relation so that the information can be easily accessed and expanded. Figure 3 shows the failure analysis process.

**① Rule-based separation of sentences:** This stage separates sentences into short clauses consisting of subject and predicate, which makes them the minimum units of a sentence. Sentences are separated using the regular expression based sentence separation rules.

**② Process-Map based Information extraction:** At this stage, the morpheme and part of speech from the short sentences separated in **step 1** are extracted and at **step 2.1.**, words are extracted from the process map as subject and predicate. After these two processes the type of that specific word is defined. If there are no exact matches with the words in the process map, partially matched words are retrieved. After that, if the extracted string from the process

map is matched with the word in the short sentence, or if the string matches the search, the corresponding string marks in the short sentence and the string is registered as the candidate of the corresponding word type. Otherwise, if the string extracted from the process map does not completely coincide with the word within the short word, the corresponding string is marked and registered as a subordinate of the candidate word type.

③ **Rule and word-position based word-type assumption & short-sentence creation:** If the type of vocabulary recognized in **step 2** corresponds to the phenomenon (function-breakdown-action), it is partitioned into separate sentences based on the vocabulary. If the vocabulary is not recognized in **step 2**, the type estimation of the unregistered word is performed. If the vocabulary is not an abbreviation and the position of the word is at the end of the sentence, it is estimated as a status term. If there are duplicated semantic meanings, the similarity degree between short sentences is calculated. If the predefined threshold is exceeded, the one with highest similarity is selected, and the short sentences with lower similarities are removed. If the word corresponding to the object or phenomenon is omitted from the natural language itself, the word restoration is performed in three steps: 1) If the word corresponding to the object of the short sentence is omitted and there is a restorable object from the previous sentence, then the object of the short sentence is restored by using the object of the previous sentence. 2) If the target word of the short sentence is omitted, but the rest of the sentences cannot be restored, then the subject of the short sentences is restored using the subject of the next sentence. 3) If the status expression of the short sentence is omitted, then the state word from the following sentence is restated.

④ **UI based word and word type modification:** The analysis of the failure report is performed automatically up to the **step 3**, and at present stage, the user can directly supplement the contents through the UI. If the user wants to change the vocabulary type of the analyzed short sentences, they can change the vocabulary by searching them from the process map and the domain vocabulary using the interface.

⑤ **Final modification by adjusting short sentence order:** This allows the user to change the order of the short sentences which are automatically analyzed up to **step 3** through the UI. The user can change the order of the selected short sentences up, down and remove the selected short sentences. In addition, the user can define new short sentences and insert them at the desired positions.

⑥ **Analyzed result insertion:** It is the step of storing the result of the analysis that is finally completed in the process map. This is done after sorting the order of short paragraphs in the order of Phenomena-Cause-Actions and then save them in the process map. In the case of a failure-related short sentence, the object and status are inserted into the failure phenomenon part of the process map, and the cause and effect relationship between the input failure phenomena is established. Also, in case of a short sentence for an action, the object and phenomenon of the short sentence are entered in the action part of the process map, and the precedence relationship between the phenomena for the action is set up. Figure 4 is an example of a failure report and an analyzed failure case. The failure report includes the name of the equipment where the failure occurred, the date and time of the failure, the failure condition, and the cause of the failure. The analyzed failure cases are represented in the list on the right side of the failure report, and they can be reconstructed in the order of the failure cause, failure phenomenon, and countermeasure Method. It can also be used as causal knowledge. The blue letter is the object of the failure and the red letter indicates the status of the facility. In addition, by providing a user interface, field experts can directly edit the objects and phenomenon of the failure and adjust the order of the failure phenomenon.

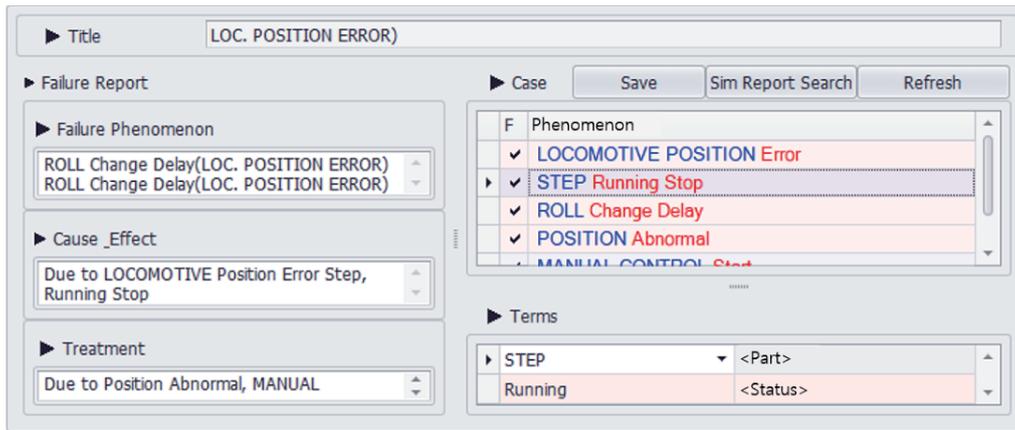


Figure 4: Sample failure report and its failure case

## 5 Alarm knowledge construction

The alarm knowledge is stored and managed in the knowledge base which is built by field experts to capture the problem of the system in real time. The expert system for alarm handling captures the failure indications from the alarm and retrieves the appropriate failure cases from the process map and provides them to field experts in predicting future failures or diagnosing current failures. The alarm knowledge constituting the system is generated after considering the relationship between failures and alarms using the user interface by field experts who have experienced a failure. Knowledge is modeled on an IF-THEN-based rule basis to represent human knowledge. The field experts generate and continuously manage the phenomena in the knowledge base which is generated by the alarm pattern. The knowledge acquisition engine also works based on the experience of the experts. When an hourly alarm is collected, the inference engine finds the appropriate rule to process the alarm from the knowledge base and derives the result. The derived results are used to find failure cases matched in the process map.

The main structure of the expert system is illustrated in Algorithm 1. The alarm list for each hour  $Alarm_i$  is used as the input for inference against the expert knowledge base  $KB$  (Line 10). The inference starts from the root rule  $R_0$ . Once the current starting rule is satisfied, all the child rules will be evaluated. Only when there is no child rule for the current starting rule will the current starting rule become the inference result (Line 18). If there is any child rule can be satisfied by the  $Alarm_i$ , the inference will continue with its child rules of this satisfied rule until no more child rules can be satisfied (Line 23 - 24). Otherwise the current starting rule is stored as one of the inference-resulted rule (Line 27). The final result of the inference  $Result$  will be a collection of rules that are satisfied by the  $Alarm_i$ . The conclusion of the each satisfied rule is in the form of the phenomenon that is stored in the process map  $PM$ . These phenomena from the conclusion will be used to find the corresponding failure case in the  $PM$  that contains the same phenomenon (Line 31-37). The failure diagnosis and prediction in the failure case can then be utilized.

### 5.1 Input case

Input cases of the expert system are the alarm lists consisting of multiple alarms and which are used for failure prediction and knowledge acquisition. The alarm system collects alarms every hour in real-time and forwards them to the expert system after processing those alarms. The alarm data consist of seven attributes such as facility ID, facility name, alarm ID, alarm name,

---

**Algorithm 1** Expert System Structure

---

```

1: LET Array  $Alarm_i \leftarrow$  Alarm list for the  $i$ th hour
2: LET rule set  $KB \leftarrow$  rule-base knowledge base
3: LET rule  $R_j \leftarrow$  the rule in the  $KB$  with the id  $j$  ( $R_0$  is a root rule)
4: LET String  $Con_j \leftarrow$  conclusion of  $R_j$ 
5: LET Array  $Result \leftarrow$  collection of satisfied rules
6: LET Object  $PM \leftarrow$  process map
7: LET case  $C_k \leftarrow$  the failure case in the  $PM$ 
8: LET String  $Phe_k \leftarrow$  the phenomenon of  $C_k$ 
9: LET Array  $Result_{PM} \leftarrow$  matched failure cases with inference result  $Result$ 
10: function INFERENCE( $Alarm_i, KB$ )
11:   if satisfied( $R_0, Alarm_i$ ) then
12:     if  $R_0$  has child rules then
13:       for all child rule  $Child$  do
14:         if satisfied( $Child, Alarm_i$ ) then
15:           if hasChild( $Child$ )=TRUE then
16:             Inference( $Child, Alarm_i$ )
17:           else
18:             push( $Result, Child$ )
19:             pop( $Result, R_0$ )
20:           end if
21:         end if
22:       end for
23:       if no child rule satisfied then
24:         push( $Result, R_0$ )
25:       end if
26:     else
27:       push( $Result, R_0$ )
28:     end if
29:   end if
30: end function
31: function RESULTMATCHING( $Result, PM$ )
32:   for all  $Con_j$  of  $R_j$  in  $Result$  do
33:     for all  $Phe_k$  in  $PM$  do
34:       if matched( $Con_j, Phe_k$ ) then
35:         push( $Result_{PM}, Phe_k$ )
36:       end if
37:     end for
38:   end for
39: end function

```

---

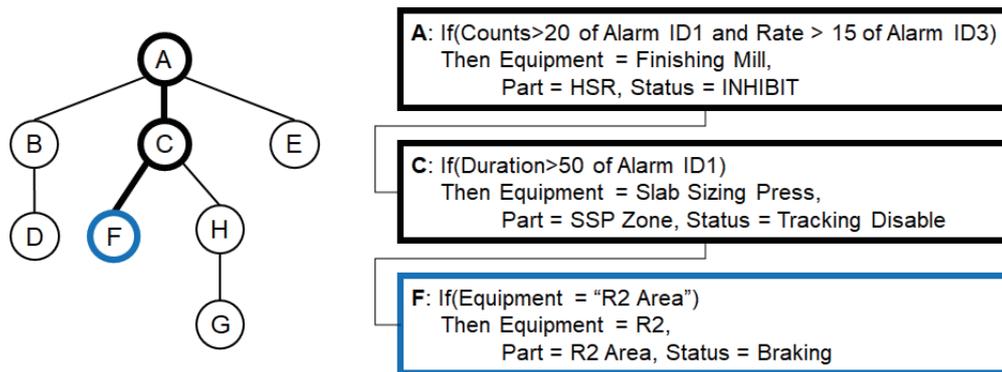


Figure 5: Knowledge base inference process

counts of alarm, the lifetime of the alarm and rate of the alarm.

## 5.2 Alarm knowledge representation

Field experts can create rules by handling alarms collected through the knowledge acquisition engine. The attributes used in the condition part of the rule are as follows: 1) Facility name: The facility that generates alarms (e.g. Finishing Mill), 2) Alarm message: The textual representation of an alarm (e.g. F3 BOT PC APC ERROR), 3) Counts: The number of times the alarm occurred in every hour (e.g. 1), 4) Duration: The duration of the alarm in every hour (e.g. 96), 5) Rate: The ratio of the alarm to one day (e.g. 26.67)

The conclusion of the rule is the same structure as the unit knowledge base in the process map. Thus, the conclusion part consists of the facilities, objects, and phenomena. The field expert who generates the rule can define the conclusion by searching and selecting the failure phenomenon in the process map. Therefore, since the inference results of the failure prediction system are mapped to the failure knowledge, a process map can be used for failure prediction. Figure 5 shows an example of a knowledge base where each rule defines the rule condition using five attributes of the alarm, and the conclusion is defined using the phenomenon which constitutes the case of the process map. For example, Rule 1 is defined as a condition with the number of occurrences (Count), alarm index (Rate), and alarm index number 3, and the conclusion of the rule is defined as Part and Status, which are phenomena of the case.

As another example, we can see in rule 2 that the conclusion of the case is based on the alarm message of alarm #1 and the conclusion of the rule is linked to the knowledge of the process map.

## 5.3 Preventive management system

The proposing preventive maintenance system monitors the current state of the facility with real-time alarms. It is also capable of conducting failure diagnosis and failure prediction via historic failure cases which is in causal-relationship format. Figure 6 represents an example of the operation of the preventive maintenance system. Four alarms are included in the alarm list collected during an hour, and the alarm list is used as the input case for the failure prediction system. Since the rule for the first alarm is stored in the knowledge base, inference engine evaluates the alarm list with the rule corresponding to the first alarm from the knowledge base and obtains the result when the rule is satisfied. The result of the rule is a failure phenomenon, and it represents that a certain phenomenon will occur at the target facility SSP zone. In order to predict possible failures from the acquired failure phenomenon, the system searches

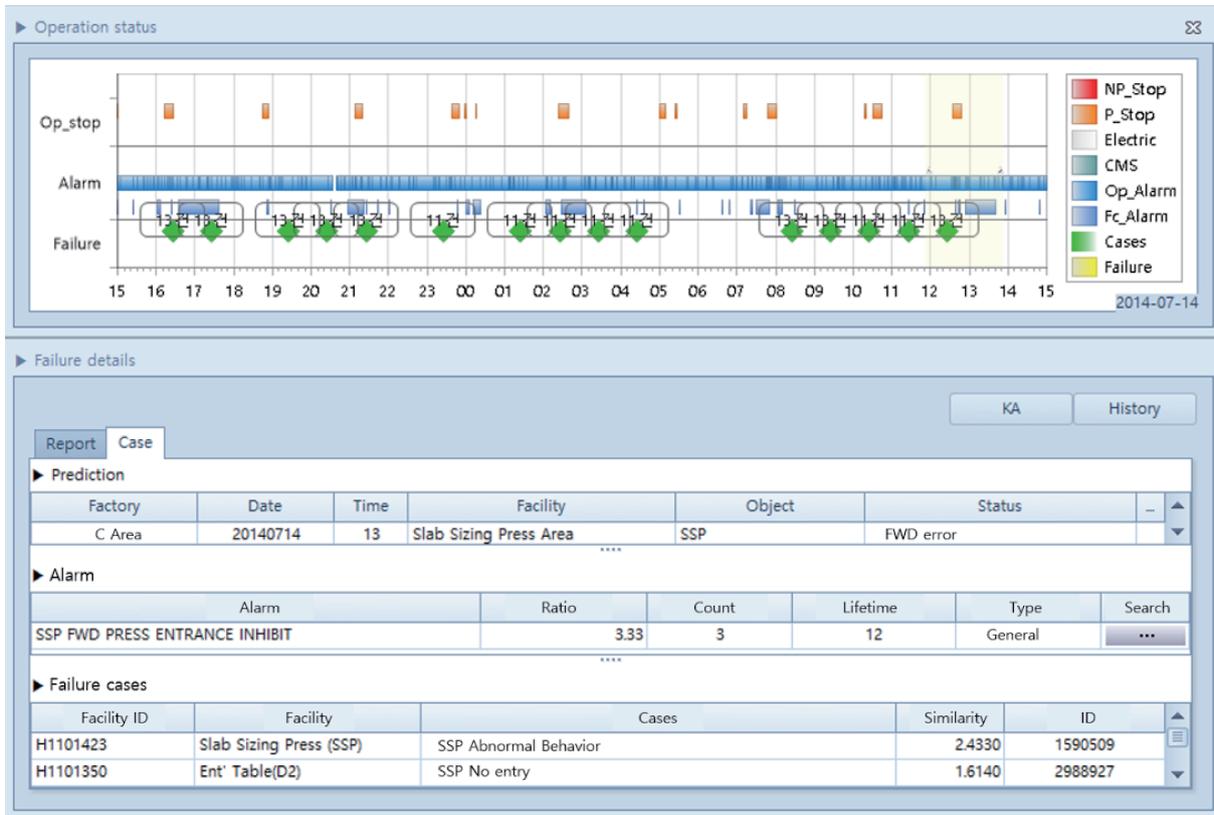


Figure 6: Preventive maintenance system interface

for failure cases which contains an equivalent failure phenomenon from the process map. The system provides the failure cases related with the facility Slab Sizing Press Area, the designated facility in the alarm, to the field expert who is monitoring the alarm. Field expert checks the phenomenon of the problematic facility and actions they can take. They are also capable of analyzing the cause of the failure by tracking the previous failure case from backward.

### 5.4 Implementation

In this paper, the example process of the proposed system is described, which currently operates in the steelworks industry. Figure 6 shows the implemented screen of the preventive maintenance system. In the alarm occurrence area located on the upper part of the screen, the occurrence status of the alarm is displayed every hour. It also displays the number of predictions for failures when the alarm occurred in the alarm occurrence area, which obtained using alarm data and failure knowledge. Users can check the details of the alarm occurrence and failure cases for the alarm via the user interface by adjusting the specific time zone. In the left bottom side of the screen, details of the alarm occurrence are displayed, including the start time and end time of the alarm, a facility which signalled the alarm, contents of the alarm, and the ratio that the individual alarm possesses. The alarms are classified as facility alarm or operation alarm. In the right bottom side of the screen, failure occurrence detail panel shows the details of the failures which occurred in the time zone that the user selected.

The failure prediction result is the inference result about the alarm that shows the SSP entrance prohibited status. The failure related alarm table shows the alarm which is primarily used during the inference process. The pseudo-failure case shows all the failure cases, including

The screenshot displays a software interface for knowledge acquisition. It features several key components:

- Alarm Load / Failure case Load:** A top-left panel showing details for a specific failure report (4959305) on 2014-05-30 at 06:22:00 PM. It lists facility information (H1104198, Table 6 (K6)) and a table of parts:
 

LV	Part	Status
1	ROT K6 MOTOR SHAFT COUPLING	Corrosion
2	RUN OUT TABLE K6 249 TABLE ROLL	Rotation disable
3	MOTOR	Connection delay
- Result(11) / Prediction(1):** A top-right panel showing a prediction table with columns for Conclusion Type, Facility, Part, and Status. The current prediction is for 'Slab Sizing Press Area' with 'SSP' status and 'No Entry'.
- Alarm(14):** A central table listing 14 alarms:
 

Facility ID	Facility	Alarm ID	Alarm	Count	Lifetime	Ratio
H1101613	R2 Area	ALM_PRC_RM_019	R2 EVEN PASS ENTRANCE INHIBIT	1	3	.83
H1101613	R2 Area	ALM_PAG_004	SDD SENSOR SYSTEM UNHEALTHY	1	3,600	1,000.00
H1101349	Slab Sizing Press Area	ALM_PRC_RM_001	SSP FWD PRESS ENTRANCE INHIBIT	2	5	1.39
H1102579	Finishing Mill	ALM_PRC_FM_091	LOOPER 6 HYD CYL STROKE EXTREME U-LIMIT	1	4	1.11
H1102579	Finishing Mill	ALM_PRC_FM_081	LOOPER 5 HYD CYL STROKE EXTREME U-LIMIT	1	4	1.11
- Rule Setup:** A bottom panel for defining rules. It includes a table for rule classification:
 

Select	Facility ID	Facility	Alarm ID	Alarm	Function	Operator	Value
<input checked="" type="checkbox"/>	H1101613	R2 Area	ALM_PR...	R2 EVEN PASS ENTRANCE IN...	Count	>=	1
<input checked="" type="checkbox"/>	H1101613	R2 Area	ALM_PA...	SDD SENSOR SYSTEM UNHEA...	Lifetime	>=	3600
<input checked="" type="checkbox"/>	H1101349	Slab Sizing Pre...	ALM_PR...	SSP FWD PRESS ENTRANCE I...	Ratio	>=	1.39
<input checked="" type="checkbox"/>	H1102579	Finishing Mill	ALM_PR...	LOOPER 6 HYD CYL STROKE ...	Lifetime	>=	1.11
<input type="checkbox"/>	H1102579	Finishing Mill	ALM_PR...	LOOPER 5 HYD CYL STROKE ...	Lifetime	>=	1.11
<input type="checkbox"/>	H1102579	Finishing Mill	ALM_PR...	LOOPER 4 HYD CYL STROKE ...	Lifetime	>=	1.11
- Rule Definition:** A panel on the right showing the logic for a rule (Rule ID: 57):
  - Condition:** ALM\_PRC\_RM\_019, Count >= 1; ALM\_PAG\_004, Lifetime >= 3600; ALM\_PRC\_RM\_001, Ratio >= 1.39
  - Conclusion:** [Conclusion ID] : 58; Part = MOTOR; Status = Connection delay

Figure 7: Knowledge acquisition from domain expert

the inference result, and it also measures, sorts, and displays the degree of similarity in order to show if the inference result is exactly included. Figure 7 represents the details of the failure case, which is displayed when the user selects a specific failure case from a failure case list in order to add a new knowledge. On the left side, the contents of the failure cases, which are generated by analyzing failure reports, sorted in the causal relationship. Each failure phenomenon is classified as failure and action, based on its attributes, which in this example, enables field experts to predict when SSP, SSP PRE-FORMING will occur. Based on those steps, the field experts can check the problem and add new failure prediction knowledge.

## 6 Evaluation

The failure prediction performance of the proposed system is evaluated through experiments, in order to prove that the failure cases of the process map indeed include causal relationships.

### 6.1 Experiment data

The data used in the experiment includes the failure reports for the failures that occurred more than once in domestic steelwork, and the alarm data collected 1 hour before and after the failure occurrence. All the alarm data are collected in real-time. In the proposed system, the data is pre-processed in the unit of 1 hour with the number of occurrences, occurred periodically, and the ratio of occurrence. From October 2012 to July 2016, a total number of 502,308 alarm data were collected. The failure report uses the failure cases built in the process map by analyzing 400 failure reports among 713, which includes the failures that occurred more than once, and are collected during the same period. The number of occurrences for the identical failures is not constant, but 4 iterated failures have occurred on average. The knowledge base of the expert system consists of 237 rules built by two field experts. For the experiment, training data and test data were used in a 6 to 4 ratio. 100 failure data and 200,923 alarm data were used as the test data.

Table 1: Top 10 satisfied rules

No.	Frequency	Rule Id	Failure Description
1	13.87%	0	Default (root) rule
2	9.98%	2	Detect the LEAK using the starting time
3	8.87%	17	Detect the BURN using lifetime
4	3.48%	201	Detect the DEFECTON using alarm id
5	2.99%	38	Detect the HUNTING using lifetime and facility id
6	2.61%	120	Detect the NO LINK using ratio and count
7	2.02%	7	Detect the NO REVERSE using alarm id and count
8	1.81%	79	Detect the SLIP using lifetime and ratio
9	1.47%	22	Detect the TRANSFORM using lifetime
10	1.03%	19	Detect the GAP using alarm id and facility id

## 6.2 Experiment method

The experiment done here consists of the following two parts. The first part shows the possibility of failure prediction method which is based on the alarm and failure knowledge, by evaluating the success rate of failure prediction with the degree where the inference results of the alarm and failure phenomenon of the failure case are mapped. The second part shows the superiority of the proposed system by comparing failure prediction accuracy of the proposed system and three previous types of research on failure prediction. Failure prediction is performed by inputting alarms into the system in the order of occurrence time, and by evaluating if the order of the reasoning results is equivalent to the order of failure phenomenon in that failure case. The evaluation is done by the following processes. For the reliability of the constructed alarm knowledge and failure cases, two field experts who are in charge of the facility monitoring at the actual steelworks evaluated the inference result of the alarm generated by the expert system. They compared the order of inference results and failure case to confirm if the cause of the actual failure and occurred failure is equivalent.

## 6.3 Experiment result

### Knowledge analysis

The ten most-frequently satisfied rules are ranked and shown in Table 1, highlighting the most frequent failure cases. The most-frequently rule was the normal (rule 0) which does not have any failure to predict. Among 400 failure type, the system found LEAK (rule 17 - If the count is larger than 9 and lifetime is more than 8 hours, Then the failure is LEAK) and BURR (rule 201 - If lifetime is less than 2 hours, Then the failure is BURR) Figure 8 shows the seasonal frequency of satisfied rules coupled with the real depth (which is their level in the decision tree), indicating failure prediction conceptual depth. Note that the root (default) rule is level 1. As can be seen in the figure, the most common rule depth is 4. Rules at this depth level includes the combination of various types of attribute sets.

### Performance evaluation of failure prediction

To evaluate the superiority of the proposing failure prediction method, we conducted a comparison with the previous failures prediction methods. The studies compared with the paper are as follows:

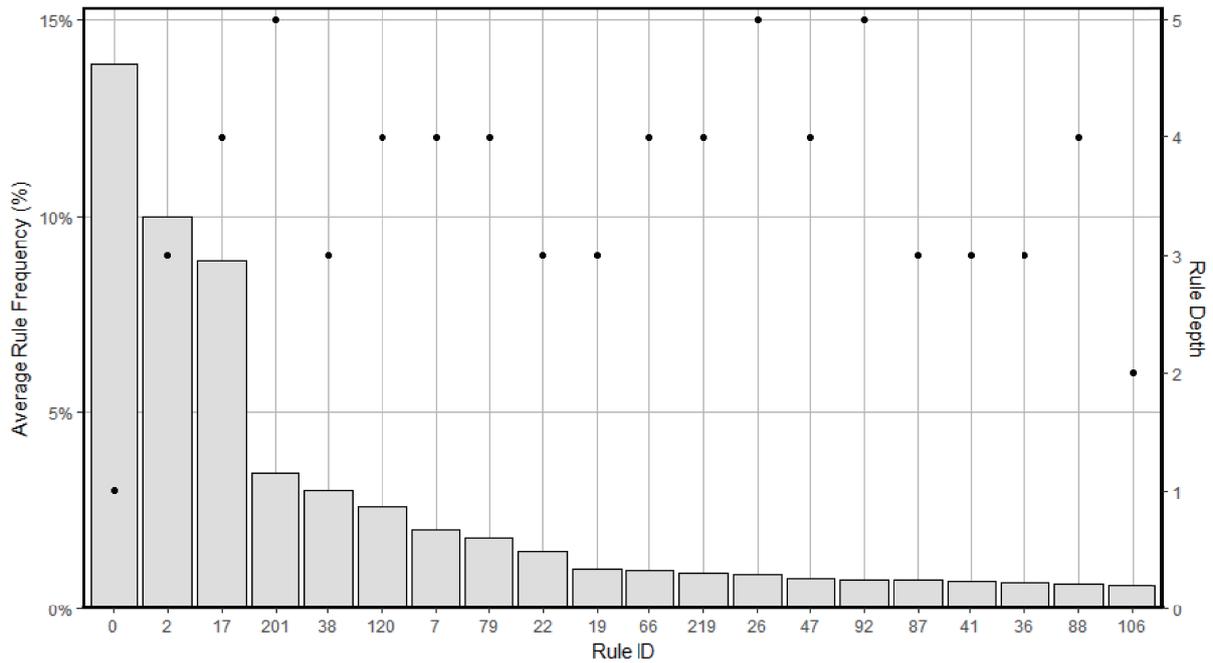


Figure 8: Inferred rule frequency and depth

Table 2: Review of failure prediction by previous failure prediction system

Author	Description	Accuracy
Santos et al. (2010)	Applied different machine learning techniques (incl. Bayesian Network, SVM, and decision tree)	81.4%
Liu and Jiang (2008)	Used particle filter with Bayesian Inference	64.2%
Chen et al. (2015)	Applied knowledge-based neural fuzzy inference	90.3%
<b>Proposed System</b>	Natural Language-based Processing Map + knowledge-based alarm prediction system	<b>95.7%</b>

- Santos et al.(2010) [13]: The authors proposed prediction method based on machine learning, in order to predict the major failures in a casting factory domain. They compared Bayesian network, SVM, and decision tree, and concluded that decision tree has the high prediction success rate.
- Liu and Jiang (2008) [9]: The authors used particle filter, which is frequently applied in signal processing, Bayesian inference, and machine learning. The research tried to predict failures with a hybrid system in the discrete-continuous composite environment.
- Chen et al.(2015) [2]: The authors used knowledge-based neural fuzzy inference for the failure prediction of turbines in wind power plant. The performance of failure prediction method proposed in this paper is evaluated by comparing with the algorithms of [13], [9], and [2] in an equivalent environment. To be specific, decision tree of [13], particle filter of [9], and neural fuzzy inference of [2] are used with the alarm data proposed in this paper. On Table 2 failure prediction accuracy between each experiment are compared.

Due to the nature of alarm data, the decision tree of [13] classified various variables into exact horizontal and vertical relations, which is not adequate for classification. As a result, the complexity of the tree grows rapidly and pruning becomes hard, which resulted in an accuracy

of 81.4%, slightly lower than the result of applying general decision tree. Since the particle filter used in [9] is based on Monte-Carlo method, which requires a massive amount of sample data, [9] showed the lowest accuracy of 64.2%. The neural-fuzzy inference method of [2] is a method which introduced learning ability of neural network into the conventional fuzzy logic method. Since it is a method which enables continuous learning by granting the learning ability to the expert knowledge-based fuzzy logic system, its key features are well utilized in the processing of complex and continuously accumulated real-time alarm data, which is reflected when it shows the similarities result recorded in its paper, 90.3%. The failure prediction success rate of the proposed method is 95.7%, which is superior to the methods used in the comparison. The reason behind this is that firstly the experts who have diverse experience in failures constructed the knowledge base directly, and secondly, the knowledge which contains real failure cases and a causal relationship is used to predict failures. Therefore, we can interpret the high accuracy as a result of utilization of high-quality knowledge which appropriately represents the actual failure cases.

## 7 Conclusion

The preventive maintenance system proposed in this paper is an effective alternative directly related to the popular smart factory for two reasons. It is based on the knowledge of experts, which greatly lowers the dependency towards human labor, and it enables effective failure prediction and diagnosis by the system. The proposed system can be utilized in various domains since it focused on the knowledge of experts which was not easily reusable before in specific domains. Proposed failure analysis system is meaningful in a perspective that it suggests new methods for knowledge sharing and generalization, which was long considered impossible. The knowledge of facilities or failures are easily obtained from manuals or reports, but the problem was that the total amount of information was enormous, and it was not easy to find exactly the information we wanted. With those reasons, the usability of failure reports went down, which made many field experts write the failure reports perfunctorily, resulting in lower low quality of reports. Such problems can be solved by improving the working environment with the help of the system. Although the technical approach of the system is meaningful to a certain extent, the system must be understood in order to be utilized in actual operation. The problem with the current failure report is that it is not easy to understand the reports for both human and systems, due to the excessive use of shortened words and specific terminology. If the human understands the working process of the system, they will write the failure reports in a way that the system could understand, which can result in a fairly accurate analysis of the system, which will reduce the dependency towards human labor. If the quality and usability of failure cases go up, the usefulness of expert system which interoperates with failure cases will also go up. In order to utilize the failure cases, the field experts will accumulate the experiential knowledge of alarm into the knowledge base regularly, and if the knowledge with high accuracy is continuously gathered to a certain extent, the system could utilize such expert knowledge to improve the accuracy of failure prediction and diagnosis with alarms. Therefore, with this process, we can overcome the disasters and human injuries, by maximizing the efficiency of preventive maintenance in the actual industrial field.

## Acknowledgements

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2017-0-01629) supervised by the IITP(Institute for Information & communications Technology Promotion). This

work was supported by the Industrial Core Technology Development Program (10049079, Develop of mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea). This work was also funded by the US Office of Naval Research grant, #GRANT12154887.

## Bibliography

- [1] Ahmed, K.; Izadi, I.; Chen, T.; Joe, D.; Burton, T. (2013); Similarity analysis of industrial alarm flood data, *IEEE Transactions on Automation Science and Engineering*, 10(2), 452-457, 2013.
- [2] Chen, B.; Matthews, P. C.; Tavner, P. J. (2015); Automated on-line fault prognosis for wind turbine pitch systems using supervisory control and data acquisition, *IET Renewable Power Generation*, 9(5), 503-513, 2015.
- [3] Cheng, Y.; Izadi, I.; Chen, T. (2013); Optimal alarm signal processing: Filter design and performance analysis, *IEEE Transactions on Automation Science and Engineering*, 10(2), 446-451, 2013.
- [4] Foong, O.; Sulaiman, S.; Rambli, D. R. B. A.; Abdullah, N. (2009); ALAP: Alarm prioritization system for oil refinery, *Proc. of the World Congress on Engineering and Computer Science*, 2, 2009.
- [5] Izadi, I.; Shah, S. L.; Shook, D. S.; Kondaveeti, S. R.; Chen, T. (2009); A framework for optimal design of alarm systems, *IFAC Proceedings Volumes*, 42(8), 651-656, 2009.
- [6] Ju, Z.; Wang, J.; Zhu, F. (2011); Named entity recognition from biomedical text using SVM; *Bioinformatics and Biomedical Engineering, (iCBBE) 2011 5th International Conference on*, 1-4, 2011.
- [7] Kang, B. H.; Kim, Y. S.; Chen, Z.; Kim, T. (2013); Detecting significant alarms using outlier detection algorithms, *Interdisciplinary Research Theory and Technology (IRRT 2013)* 1-8, 2013.
- [8] Langone, R.; Alzate, C.; Bey-Temsamani, A.; Suykens, J. A. (2014); Alarm prediction in industrial machines using autoregressive LS-SVM models, *Computational Intelligence and Data Mining (CIDM), 2014 IEEE Symposium on*, 359-364, 2014.
- [9] Liu, Y.; Jiang, J. (2008); Fault diagnosis and prediction of hybrid system based on particle filter algorithm, *Automation and Logistics, 2008. ICAL 2008. IEEE International Conference on*, 1491-1495, 2008.
- [10] Mohapatra, H.; Jain, S.; Chakrabarti, S. (2013); Joint Bootstrapping of Corpus Annotations and Entity Types, *EMNLP*, 436-446, 2013.
- [11] Morwal, S.; Jahan, N.; Chopra, D. (2012); Named entity recognition using hidden Markov model (HMM), *International Journal on Natural Language Computing (IJNLC)*, 1(4), 15-23, 2012.
- [12] Orair, G. H.; Teixeira, C. H.; Meira Jr, W.; Wang, Y.; Parthasarathy, S. (2010); Distance-based outlier detection: consolidation and renewed bearing, *Proceedings of the VLDB Endowment*, 3(1-2), 1469-1480, 2010.

- [13] Santos, I.; Nieves, J.; Bringas, P. G. (2010); Enhancing fault prediction on automatic foundry processes, *World Automation Congress (WAC)*, 1-6, 2010.
- [14] Sawsaa, A.; Lu, J. (2011); Extracting information science concepts based on jape regular expression, *WORLDCOMP'11The 2011 World Congress in Computer Science, Computer Engineering, and Applied Computing*, 18-21, 2011.
- [15] Zhao, W.; Bai, X.; Wang, W.; Ding, J. (2005); A novel alarm processing and fault diagnosis expert system based on BNF rules, *Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES*, 1-6, 2005.
- [16] Zhu, J.; Shu, Y.; Zhao, J.; Yang, F. (2014); A dynamic alarm management strategy for chemical process transitions, *Journal of Loss Prevention in the Process industries*, 30, 207-218., 2014

# Decoupling 5G Network Control: Centralised Coordination and Distributed Adaptation

S.J. Lin, J.G. Yu, W. Ni, R.P. Liu

**Shangjing Lin\***, Jianguo Yu

Beijing University of Posts and Telecommunications  
Beijing, China, 100876  
linshangjing@bupt.edu.cn, yujg@bupt.edu.cn  
\*Corresponding author: linshangjing@bupt.edu.cn

**Wei Ni**

Data 61, CSIRO  
Australia, 2122  
Wei.Ni@data61.csiro.au

**Ren Ping Liu**

University of Technology, Sydney  
Australia, 2122  
Ren.Liu@uts.au

**Abstract:** Fifth generation mobile networks (5G) will be featured by miniaturised cells and massive dense deployment. Traditional centralised network control cannot adapt to high signalling delay, and is therefore not scalable for future 5G networks. To address this issue, we adopt the software-defined networking (SDN) approach of decoupled network control and data transmission. In particular, delay-sensitive interference suppression for data transmission is decoupled from delay-tolerant topology control and base station coordination. This substantially alleviates the requirement of network control on delay and complexity, hence simplifying 5G control plane design, reducing signalling overhead, and enhancing network scalability. Case studies show that our decoupled network control is effective for timely interference mitigation and reliable topology management. The stability and scalability of our approach are also demonstrated.

**Keywords:** fifth generation mobile network (5G); network control; protocol stack; standardisation.

## 1 Introduction

Fifth generation mobile networks (5G) will be featured by miniaturised cells and massive dense deployment [3]. The features are driven by the goal of a thousand times increased network capacity, compared to the current 4G systems [13]. They are further triggering a paradigm shift of network architecture and operation. Privately owned/installed plug-and-play base stations and individually subscribed backhaul connections are promising in future 5G deployments.

Fig. 1 illustrates the promising future 5G architecture, where plug-and-play 5G base stations are connected to the Servicing Gateways (S-GWs) of the 5G core network for interference coordination and mobility control. The backhaul connections between the base stations and the S-GW can be heterogeneous, subscribed to different Internet Service Providers (ISPs) (like existing household WiFi). Such architecture is able to support the massive dense deployment of 5G base stations, using existing network infrastructure and saving deployment cost.

Meanwhile, critical challenges are arising in the control of the heterogeneous 5G networks. Particularly, high and substantially unbalanced delay of hundreds of milliseconds may occur on the subscribed backhaul connections between the base stations and the S-GW [10]. Close and

timely network control that has been necessary in existing cellular systems (with dedicated transport networks) for interference mitigation [1] becomes impossible in the heterogeneous 5G systems. However, timely control will remain crucial in drastically changing wireless environments where interference levels can change instantly. Fast power control at intervals of milliseconds will still be necessary to mitigate interference in future 5G systems.

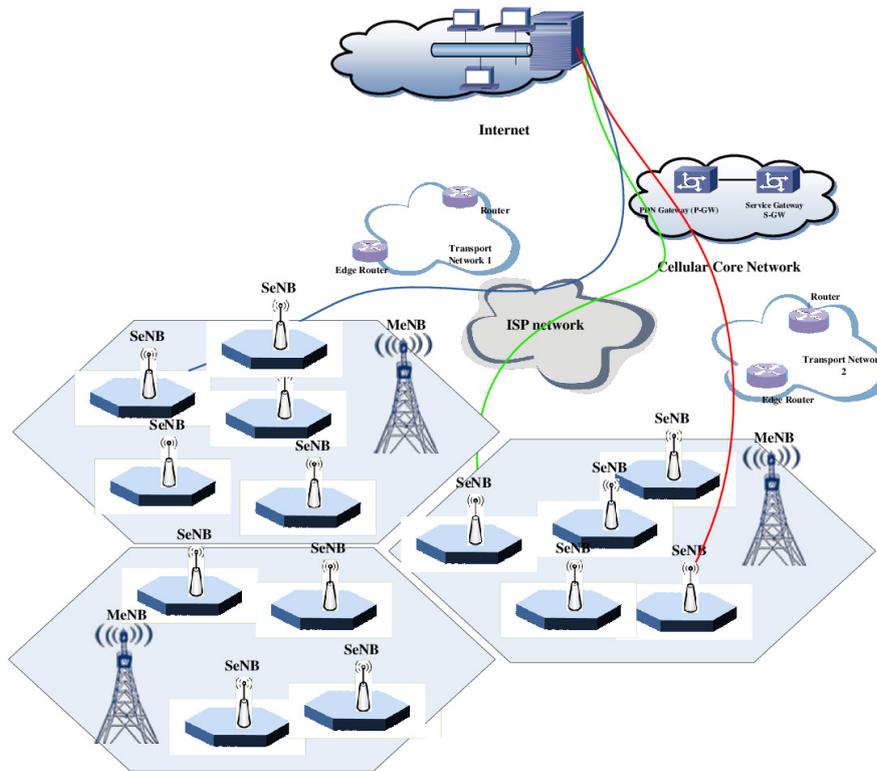


Figure 1: An illustration on the heterogeneous 5G architecture, where some base stations are connected to the S-GW via cellular infrastructure, as highlighted by the red path, while most base stations are connected to the S-GW through various ISP networks, as indicated by the green and blue paths. Centralised control signalling travels via these colour-coded paths.

Earlier, the concept “software-defined networking (SDN)” was proposed to split control and data in wired networks, which can simplify network control and reduce cost [5]. The SDN concept was experimentally demonstrated effective for its application to wired IP/Ethernet transport networks of mobile systems [11]. Recently, the SDN idea of centralised control and distributed data delivery has been extended for the control of radio access networks (RANs) [4, 6–9, 11]. Nevertheless, all the extensions would experience the above issue of delayed control if applied to the heterogeneous 5G networks, and fail to suppress interference in a timely manner.

One of the SDN extensions to wireless applications was big base station abstraction [4], where a controller models base stations as a dimension of resources, and assigns them and powers to traffic flows. Unfortunately, the abstraction was based on an oversimplified assumption of independence between base stations and frequencies. As a consequence, interference, which can substantially differ in practice when different pairs of base stations are allocated with the same frequency, was treated as static.

Other extensions of the SDN centralised network control were on software-defined real-time pairing of antennas and base stations, adapting to traffic demand and interference [7–9]. They are promising for a local 5G deployment with a dedicated (operator controlled) transport network.

However, the extensions are unsuitable for the heterogeneous 5G architecture involving ISP transport networks and incurring delayed control. They are also unsuitable for large and densely deployed systems, where centralised optimisations are computationally prohibitive and limit the network scalability.

In this article, we propose to decouple 5G network control and distribute control computations, so that the critical issues of delayed control and limited scalability can be tackled. In particular, delay-sensitive interference suppression for data transmission is decoupled from delay-tolerant topology control and base station coordination. The interference suppression can be achieved quickly through distributed adaptation of base stations. The topology and frequencies are progressively coordinated through a centralised SDN controller. As a result, the requirement of network control on delay and complexity can be substantially alleviated, simplifying 5G control plane design, reducing signalling overhead, and enhancing network scalability. Case studies show that our decoupled network control is effective for timely interference mitigation and reliable topology management. The stability and scalability of our approach are also demonstrated.

## 2 Evolutionary view of 5G network control

As discussed in Section 1, centralised network control faces critical challenges of severely delayed control and limited scalability in a heterogeneous 5G network depicted in Fig. 1. Decentralisation of network control is the way to eliminate the challenges, and therefore of practical value.

We note that a 5G base station will not only be a signal generation and processing entity, but it will also have self-organising capabilities of running a variety of operations specified by software. In this sense, 5G base stations will be equipped with a certain level of intelligence, and capable of automation and local decision-makings. The decisions that a base station can make can include switching from severely interfered (crowded) wireless channels to less interfered channels, as well as adjusting its own transmit power.

Given the intelligence of individual base stations, the entire network can be visualised as an evolution process driven by changing traffic demand, wireless channel conditions and interference levels. In particular, the network topology and the radio resources assigned with the topology evolve along with those changes. The network topology here describes clusters of cells in the 5G network, where cells belonging to a cluster reuse the same frequency assigned to the cluster. Each individual of the base stations drives the evolution of the topology by forming clusters and switching between the clusters (i.e., from severely interfered clusters to less interfered clusters). The entire population of the base stations evolves towards stable topology and balanced interference.

Fig. 2 shows an example of the network evolution, where only four base stations and two clusters are considered for illustration convenience. The evolution consists of two stages. In the first stage, the base stations within each cluster quickly and interactively adjust their transmit power through distributed automation to balance interference. In the second stage, the base stations choose to stay in the current clusters or switch out to less interfered clusters. This is a speciation stage in the evolution. It requires a brief idea of the entire base station population, and therefore needs centralised assistance. The two stages take place in an alternating manner, until no base stations want to switch clusters and the system is stabilised.

Such an evolutionary view of network control allows for the decomposition of the network control into centralised and distributed parts. In light of this, a decoupled, hybrid protocol stack for 5G network control is developed, as will be described in Section 3.

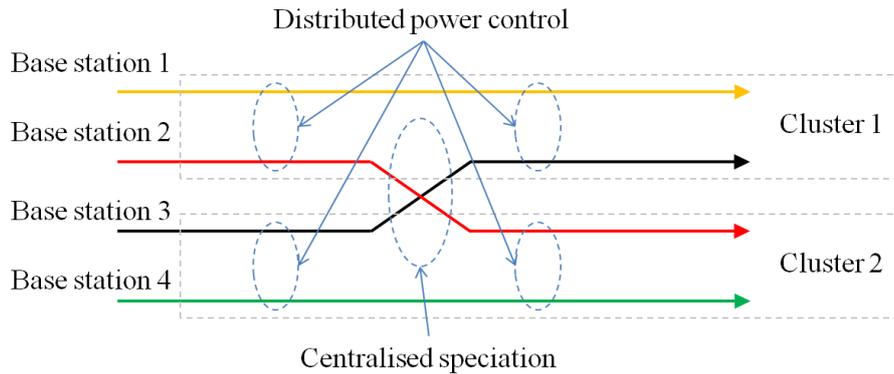


Figure 2: An illustration on the evolution modelling of 5G networks, where colour-coded arrow lines represent different base stations.

It is worth pointing out that IEEE 802.11 access point (AP) can also be treated as rational with self-organising capabilities. However, the APs cannot interact with each other, and therefore are unable to suppress interference. What they can do is to detect and select radio channels with negligible interference on the installation and initialisation.

### 3 Decoupled, hybrid protocol stack for 5G network control

The new protocol stack that we propose for 5G network control consists of two layers, namely, upper layer and lower layer. The two layers are designed to decouple the delay-tolerant and delay-sensitive aspects of network control, driven by the evolutionary view of 5G network established in the last section.

- a. The upper layer of the protocol stack is a centralised, delay-tolerant network control, which resides in both base stations and the network controller, tolerates hundreds of milliseconds delay over transport networks, and takes the responsibility of slow-changing topology control and resource management.
- b. The lower layer of the protocol stack is a distributed, delay-sensitive network control, which resides in the base stations, and is implemented by the automation of 5G devices to combat fast-changing wireless channels and interference at an interval of milliseconds.

The two layers are interactive, distributing computations and balancing the complexity. As a result, signalling overhead can be reduced and network scalability can be enhanced.

#### 3.1 Upper layer: Centralised coordination

The upper layer of the new 5G protocol stack is to manage the network topology and frequency, so as to balance interference and satisfy traffic demands. The upper layer is interconnected in a star topology from the centralised network controller to individual base stations.

The network topology can be described by the clusters of the cells in the network, denoted by  $\mathcal{C}_n$  ( $n = 1, \dots, N$ ).  $N$  is the number of clusters. Base stations belonging to each of the clusters reuse the same frequency. The base stations may need to switching between clusters (i.e., from the severely interfered clusters to the less interfered) to alleviate interference and meet traffic demand. This causes the dynamics of the network.

The upper layer of the new protocol stack models the network dynamics as an evolutionary (game) process. The recent SDN concept can be extended as such that the centralised network

controller provides guidance to the base stations, assisting them to make local cluster-switching decisions. The network controller can also steer the network evolution towards stabilised topology and balanced traffic. The reason for allowing local (distributed) decision-makings is that the globally optimal cluster-switching decisions are computationally and practically impossible in densely deployed 5G small cells, as discussed in Section 1.

The centralised network controller can be designed to distribute the available 5G spectrum between the clusters  $\mathcal{C}_1, \dots, \mathcal{C}_N$  in such a fair fashion that the total achievable data rate of each cluster is equally proportional to the request. Methods, such as Shapely value of cooperative games [12], can be used to distribute the frequencies.

The centralised controller can then evaluate the current clustering by measuring the probability distribution of the achievable data rate  $r_i$  of every cell  $i$  in a cluster  $\mathcal{C}_n$  ( $i \in \mathcal{C}_n$ ). The distribution function is denoted by  $F_n$ . The controller further evaluates the probability distribution with regards to all the clusters. The distribution function is denoted by  $F$ .

For the clusters whose distribution functions  $F_n$  are left to  $F$  (i.e., the clusters have lower achievable data rates), the centralised network controller specifies probabilities for their base stations to switch to other less interfered clusters whose distribution functions are right to  $F$ .

The probabilities can be specified, depending on the ranking of the base stations within their current clusters and the distances between the distributions of the current clusters and the target clusters. The base stations will have higher probabilities to switch out of a cluster  $\mathcal{C}_i$ , if  $F_i$  is further away from left to  $F$  (i.e.,  $\mathcal{C}_i$  is crowded). They will have higher probabilities to switch into  $\mathcal{C}_j$ , if  $F_j$  is further away from right to  $F$ .  $i \neq j$ . Within the cluster  $\mathcal{C}_i$ , the lower data rate that a base station can achieve, the probability will be higher for the base station to switch out of the cluster, since the base station suffers more severe interference under the current clustering.

The first step of the clustering review is to specify the ratio of the femtocells that need to switch clusters. This is an evolutionary process. The ratio of the femtocells that need to switch out of  $\mathcal{D}_m$  can be specified by

$$\begin{aligned} \dot{p}_m &= p_m \times \|f_m(x) - \bar{f}(x)\|_{\mathcal{D}} \\ &= \frac{p_m}{\sum_{\forall m} Q_m} \text{sign}\left(\mathbb{E}_m(x) - \mathbb{E}(x)\right) \times \\ &\quad \underbrace{\left(1 - \int_0^{\infty} \min\left(f_m(x), \bar{f}(x)\right) dx\right)}_{Q_m}, \end{aligned} \quad (1)$$

where  $p_m$  is the ratio of femtocells in  $\mathcal{D}_m$ ;  $\|\cdot\|_{\mathcal{D}}$  denotes the statistical difference between two clusters, which can be calculated by  $Q_m / \sum_{\forall m} Q_m$ ;  $\text{sign}(\cdot)$  is the sign function.

$Q_m$  denotes the under-braced part, in which the integration calculates the area of  $f_m(\cdot)$  that is not overlapped with  $f(\cdot)$ .  $\mathbb{E}_m(\cdot)$  and  $\mathbb{E}(\cdot)$  are the means associated with the two PDFs. The sign function indicates that the calculated non-overlapping area is to the left of  $f(\cdot)$ , if the sign is negative; or to the right of  $f(\cdot)$ , otherwise.  $Q_m$  is normalized by  $1 / \sum_{\forall m} Q_m$ .

Such design of  $\dot{p}_m$  allows the evolution process of clustering towards reducing the distributional differences between the clusters. Clusters that are below average (i.e., their PDFs are to the left of  $f(\cdot)$ ) will have their PDFs adjusted rightwards by having some of their femtocells switch to other clusters. Specifically,  $\dot{p}_m < 0$  means that  $-\dot{p}_m$  of  $|\mathcal{D}_m|$  femtocells should switch out of  $\mathcal{D}_m$ , and  $\dot{p}_m > 0$  means that  $\dot{p}_m$  of  $|\mathcal{D}_m|$  femtocells should switch into  $\mathcal{D}_m$ . For a femtocell that is to switch out of  $\mathcal{D}_m$  ( $\dot{p}_m < 0$ ), it switches into  $\mathcal{D}_j$  at the probability of  $\dot{p}_j / \sum_{\forall i: \dot{p}_i > 0} \bar{p}_i$  ( $j \neq m$ ).

Next, the second step of the clustering review and adjustment is for every femtocell  $i \in \mathcal{D}_m$  to specify its own probability to switch clusters, denoted by  $\dot{p}_{m,i}$ . This is based on  $\dot{p}_m$  and the

satisfaction level of femtocell  $i$ . This step differentiates the femtocells within the same cluster in terms of switching probability.

To do this, we develop the following design.

$$\dot{p}_{m,i} = \dot{p}_m \frac{\lambda_m e^{\lambda_m}}{e^{\lambda_m} - 1} \exp\left(-\lambda_m F_m(x_i)\right) \quad (2)$$

where  $x_i = R_i^{\text{gnt}}/R_i^{\text{req}}$ , and  $F_m(\cdot)$  is the cumulative distribution function (CDF) of the satisfaction levels within  $\mathcal{D}_m$ .  $\lambda_m$  is a parameter that can be adjusted so that  $\dot{p}_{m,i} \leq 1$ . The granted data rate  $R_i^{\text{gnt}} = \phi_m \omega_c \log_2(1 + \gamma_i^*)$ ,  $i \in \mathcal{D}_m$ .

The rationale for this design is that (2) does not change the overall clusterwise switching probability of the cluster, i.e.,  $\dot{p}_m$ . Specifically, we have

$$\begin{aligned} \lim_{|\mathcal{D}_m| \rightarrow \infty} \frac{1}{|\mathcal{D}_m|} \sum_{i \in \mathcal{D}_m} \dot{p}_{m,i} &= \int_0^\infty \dot{p}_{m,i} f_m(x) dx \\ &= \dot{p}_m \frac{\lambda e^{\lambda_m}}{e^{\lambda_m} - 1} \int_0^\infty \exp\left(-\lambda_m F_m(x)\right) f_m(x) dx \\ &= \dot{p}_m \frac{e^{\lambda_m}}{e^{\lambda_m} - 1} \int_0^{\lambda_m} e^{-u} du = \dot{p}_m. \end{aligned} \quad (3)$$

Meanwhile, individual femtocells in the cluster can have different switching probabilities, depending on their individual satisfaction levels. As such, femtocells that receive excessive interference in crowded clusters get high possibilities to switch to other less interfering clusters.

Note that the second step of creating differential switching probabilities in a cluster is new. Individual femtocells' statuses are leveraged without invalidating the evolution of the entire population (which was specified in the first step). In contrast, conventional evolutionary games are focused on the entire population, ignoring the differences between individuals.

As an effect of the two-step design, the satisfaction distributions of different clusters converge through the evolution of clustering. The average satisfaction level of all the femtocells gradually grows until the convergence.

After the second step, each femtocell switches its cluster based on the probability specified in (2). The new  $M$  clusters,  $\{\mathcal{D}_1, \dots, \mathcal{D}_M\}$ , will be considered in the next round of distributed power adaptation (as described in Section 3.2).

### 3.2 Lower layer: Distributed adaptation

The lower layer of the new 5G protocol stack describes a distributed automation process of every base station in terms of transmit parameters, adapting to the network topology change driven by the upper layer. Specifically, every base station, say base station  $i$  in the  $n$ th cluster ( $i \in \mathcal{C}_n$ ), collects interference measurement reported by its mobile terminal and independently updates its transmit power  $P_i$ . The base station does so iteratively, until the transmit power is stabilised. The cluster also becomes stabilised.

A well designed utility function, with which the transmit power is independently updated in response to the measured interference, is the key to the lower layer of the protocol stack. The function will lead individual base stations to stabilise and converge in a distributed manner to the stable point of an entire cluster. This requires the function to have a convex/concave structure with respect to all the transmit powers of the cluster. Meanwhile, the data rate of each base station should not be compromised at the stable point for the convexity/concavity.

An example of such function is given by

$$U_i(P_i, \mathbf{P}_{-i}) = \arctan\left(\frac{\gamma_i(P_i, \mathbf{P}_{-i})}{\Gamma_i} \alpha_n\right) - \theta_n P_i, \quad i \in \mathcal{C}_n \quad (4)$$

where  $\mathbf{P}_{-i}$  collects  $P_j$  for  $j \in \mathcal{C}_n$  and  $j \neq i$ ;  $\gamma_i(P_i, \mathbf{P}_{-i})$  is the signal-to-interference-plus-noise ratio (SINR) of the terminal served by cell  $i$ ;  $\Gamma_i$  is the target SINR that is required to meet the traffic demand of the terminal;  $\alpha_n$  is a predefined coefficient to adjust the speed of the stabilisation of  $P_i$  in cluster  $\mathcal{C}_n$ ;  $\theta_n$  is a predefined coefficient to adjust the weights of the two terms in the function.

$U_i(P_i, \mathbf{P}_{-i})$  is designed to be maximized at each base station  $i \in \mathcal{C}_n$ . The first term at the right-hand side of the function defines the utility which drives the achieve SINR towards the target. This is through increasing the transmit power  $P_i$ . The second term at the right-hand side of the function is the cost/penalty to achieve the utility. It keeps the growth of  $P_i$  under control and prevents generating excessive interference to other cells.

The function can ensure the stabilisation of each transmit power, e.g.,  $P_i$ . This is due to the concave structure of the function with respect to all the  $P_k$ 's with  $k \in \mathcal{C}_n$  [2]. The concavity can be rigorously proved by evaluating the Hessian matrix of the function and confirming that the Hessian matrix is negative definite with carefully selected  $\alpha_n$  and  $\theta_n$ . As a result, the function has a unique stable point that is the global optimum of the function, as given by

$$P_i^* = \min \left( \frac{I_{-i}(\mathbf{P}_{-i})\Gamma_i}{\alpha_n g_i} \sqrt{\frac{1-\theta_n}{\theta_n}}, P_{\max} \right), \quad (5)$$

where  $I_{-i}(\mathbf{P}_{-i})$  is the total interference from other cells in cluster  $\mathcal{C}_n$  to cell  $i$ ,  $g_i$  is the channel gain of base station  $i$  to its terminal, and  $P_{\max}$  is the maximum transmit power of the base station. All the cells will be stabilised at the point.

The concavity is also crucial to ensure all the cells to be stabilised at the stable point through distributed automation [12]. This is due to the fact that, on each plane cutting through the concave structure, the intersecting curve is concave. In other words,  $U_i(P_i, \mathbf{P}_{-i})$  is concave with respect to  $P_i$  for any given  $\mathbf{P}_{-i}$ . The base station  $i$  is able to independently update the transmit power  $P_i$  towards increasing the utility, based on measured interference.

Let  $\tilde{I}_{-i}(k)$  denote the  $k$ -th measurement result of interference ( $k = 1, 2, \dots$ ). It can be measured by the mobile terminal on the air interface, and reported to the base station through the virtual interface "I\_d".  $\tilde{I}_{-i}(k)$  can replace  $I_{-i}(\mathbf{P}_{-i})$  in (5) and update  $P_i^*(k+1)$  in practical distributed implementations, pushing all the base stations to converge to and stay at the unique stable point of (5). As a result, the requirement of signalling exchange between the cells can be eliminated, as well as the centralised coordination of the upper layer of the protocol stack.

Other utility functions with convex/concave structures can also be designed. Stable points with increased convergent SINRs are expected with proper designs of the functions.

### 3.3 Interactions between the two layers

The two layers of the new 5G protocol stack are interactive, as illustrated in Fig. 3. On one hand, the distributed automation of the lower layer depends on the network topology and resource management outcome of the centralised control of the upper layer. On the other hand, the centralised network control of the upper layer needs to operate on stabilised systems after the distributed automation of the lower layer, so that the network topology and resources can be adapted and updated.

The interaction between the two layers is facilitated by the interface "I\_c/d" at each base station. In the downstream direction, the evolving topology that was updated at the upper layer and carried by "I\_c" is passed down to the lower layer through "I\_c/d" to activate distributed power control at the lower layer. In the reverse direction, transmit powers stabilised at the lower layer are handed over to the upper layer through "I\_c/d", and will gather at the centralised network controller through "I\_c" to trigger a new round of topology management.

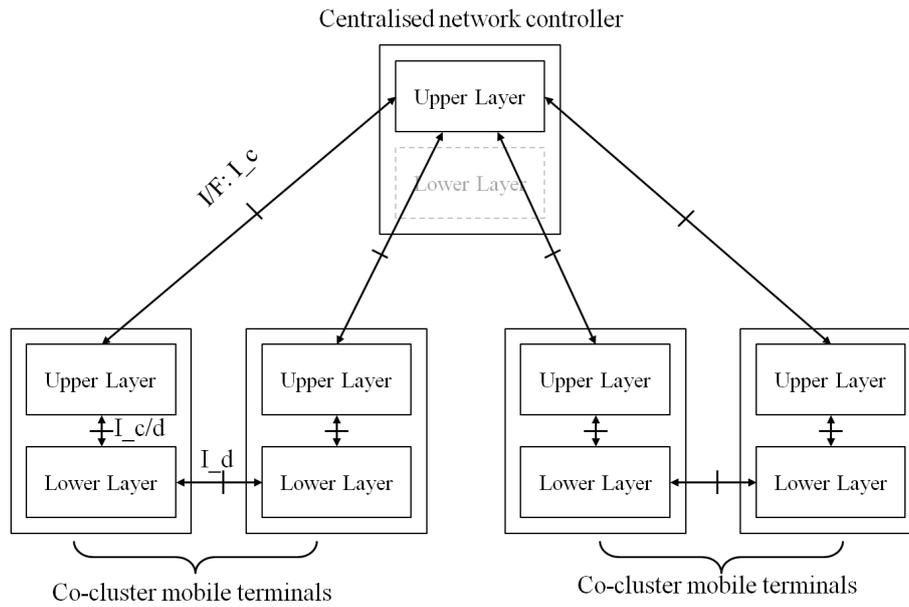


Figure 3: An illustration on the decoupled 5G control protocol stack, where the upper layer of the protocol stack is interconnected in a centralised manner through the interface (I/F) "I<sub>c</sub>" between base stations and the centralised network controller, the lower layer of the protocol stack behaves in a distributed manner through a virtual interface "I<sub>d</sub>", and the two layers are interactive through the interface "I<sub>c/d</sub>" with every base station

The star topology of the upper layer also provides signalling bearers via the interface "I<sub>c</sub>" to carry the probabilities to the base stations. On the receipt of the probabilities through the upper layer, each base station independently decides whether to switch clusters and which cluster to switch to.

As a result of the centralised upper layer, cells migrate from crowded clusters to less interfered clusters. The distributions of data rate converge across the clusters, as well as getting less spanned. Fairness is improved between the clusters, and also within a cluster. Moreover, the evolution design of the upper layer eliminates the requirement of signalling exchange between clusters, and tolerates the great level of uncertainty in large practical wireless networks. The design also distributes the complexity of clustering decision-makings between the centralised controller and individual base stations. The scalability of the network is therefore enhanced.

## 4 Performance studies

Extensive MATLAB simulations have been conducted to evaluate the hybrid protocol stack for 5G network control. The case where the evaluation is carried out is a heterogeneous 5G network covering a geographical area of  $500^2\pi\text{m}^2$ , where there are 30 available frequency channels and each channel has a bandwidth of 180 KHz.

Consider that the privately owned/installed plug-and-play base stations are most likely to be indoors and operate at a low power level. We assume that the maximum transmit power of each base station is 20 dBm. We also assume a wall penetration loss of 10 dB when the radio signals of a 5G cell leak out from indoors to outdoors and become interference to other cells. The path loss exponent is 3.7. The receive noise at the mobile terminals is set to be  $-174$  dbm/Hz. The traffic demand is assumed to be 5.0 Mb/s per mobile terminal.

Fig. 4 shows the overall throughput of the 5G network with the increasing number of cells, where the number of clusters grows from one to five. Apart from our hybrid approach, we also plot the pure centralised SDN network control proposed in [6], where clusters are constructed using graph theory to maximize the distance between the clusters, and the base stations transmit a fixed power of 20 dBm since fast power control at milliseconds intervals is impossible due to centralised and severely delayed control. The figure demonstrates that, in general, the hybrid network control provides higher throughput than the pure centralised control. When the evolutionary clustering is on, the hybrid network control adaptively decides the number of clusters. The throughput under the hybrid network control operates as the topmost of all the five solid curves, which is far beyond the throughput of the centralised approach.

A close look at each of the five solid curves shows that for every given number of clusters, the throughput starts by growing, due to the increased traffic demand resulting from the increased cell density. After reaching a peak, the throughput rapidly drops. The reason is because the network is so dense, the clusters are becoming very crowded, and the interference becomes excessively large. Distributed automation of power control or even switching between clusters cannot help to suppress the interference. The excessive interference not only offsets the increased throughput resulting from increased demand, but also adversely affects and further reduces the throughput. In this sense, the peak corresponds to the case where the network density and interference are balanced and the throughput is maximized given the number of clusters.

A joint look across all the five solid curves in Fig. 4 shows that increasing the cluster number is able to increase the peak throughput, as well as the number of cells achieving the peak. Of course, clustering can help alleviate interference by reducing the number of cells using the same frequency, as shown in the figure where the number of cells that can be supported increases with the number of clusters. On the other hand, clustering also decreases the bandwidth that every cell can use, as depicted in the figure where the throughput decreases with the increasing number of clusters in the presence of a small number of cells (i.e.,  $N \leq 50$ ). The conclusion that we draw is that the significantly reduced interference due to the evolutionary clustering is able to compensate for the reduced bandwidth, when the cell density and interference are balanced. As a result of this, the peak throughput keeps growing with the cluster number.

Fig. 5 plots the average transmit power required to achieve the throughput performance in the previous figure, with respect to the increasing number of cells. The transmit power is the average of all the base stations' after they are stabilised from the distributed power control automation. We can see that, in general, the stabilised transmit power decreases with the increase of the cluster numbers. This is because clustering helps alleviate interference, which in turn reduces the requirement of transmit power. When the evolutionary clustering of our hybrid network control is on, the average transmit power is as shown as the highlighted black zigzag curve in the figure. This is because the number of clusters is adaptively adjusted to maximize the throughput, when the evolutionary clustering is on. The number of clusters may increase, as the number of active cells grows. This can lead to a sudden drop of the transmit power, partly because of the alleviated interference and partly because of the reduced bandwidth per cell.

It is also interesting to notice that the peaks on the highlighted black curve grows slowly with the number of cells. This is consistent with Fig. 4, indicating that increased transmit powers are necessary to improve the achievable throughput. The sudden drop of the transmit power from the peaks also indicates that when interference becomes intensive, releasing part of the frequencies for a new cluster can substantially relieve the interference. As a result, much lower transmit powers are required to achieve the same throughput. In other words, the peaks are where the base stations should trade out their bandwidth for relieved interference.

Fig. 6 shows the convergence of the average throughput of each cluster along with evolution over time, where a single evolution process of 5 clusters is recorded. It demonstrates the conver-

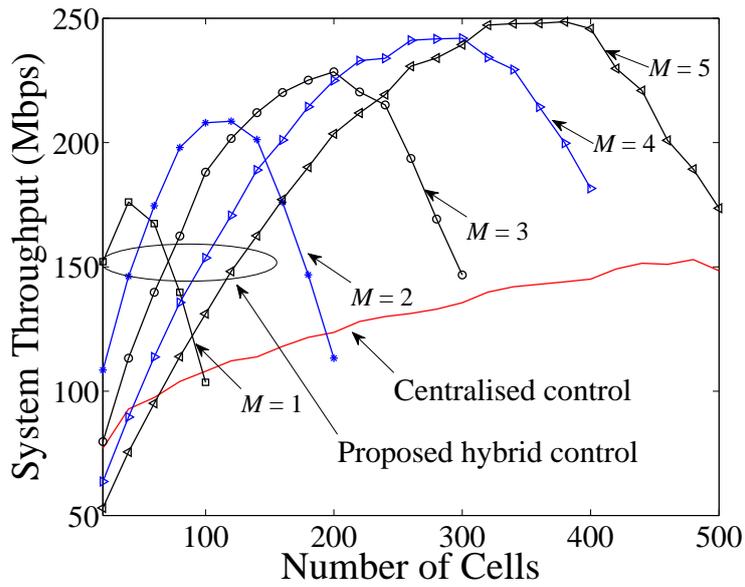


Figure 4: Overall system throughput versus the number of cells, where the number of clusters, denoted by  $N$ , grows from 1 to 5.

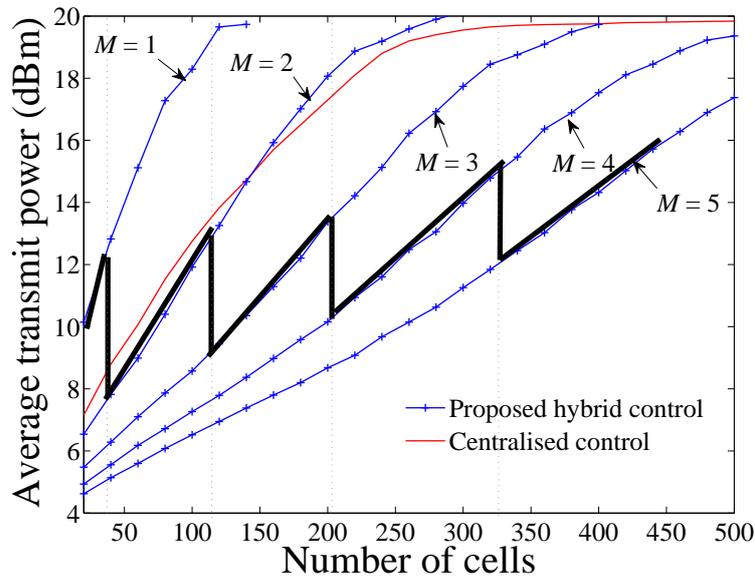


Figure 5: Average transmit power per base station versus the number of cells, where the black curve is the actual required average transmit power when the evolutionary clustering of our hybrid network control is on, and is plotted by picking the crossing points of any two adjacent throughput curves in Fig. 4.

gence of the clusters in terms of throughput. The largest difference among the clusters decreases from 0.8 Mbps to 0.1 Mbps after 320 evolution stages (i.e., 6.4 seconds), and the difference diminishes by 500 evolution stages (i.e., 10 seconds). We also see that most of the clusters can

fast converge to balancing their throughput. Specifically, clusters 1 to 4 converge by 200 stages. Cluster 5 does not converge as fast as the others do, because of its small initial size (as will be shown in Fig. 7). However, the convergent tendency of the cluster is clear, as shown in the figure.

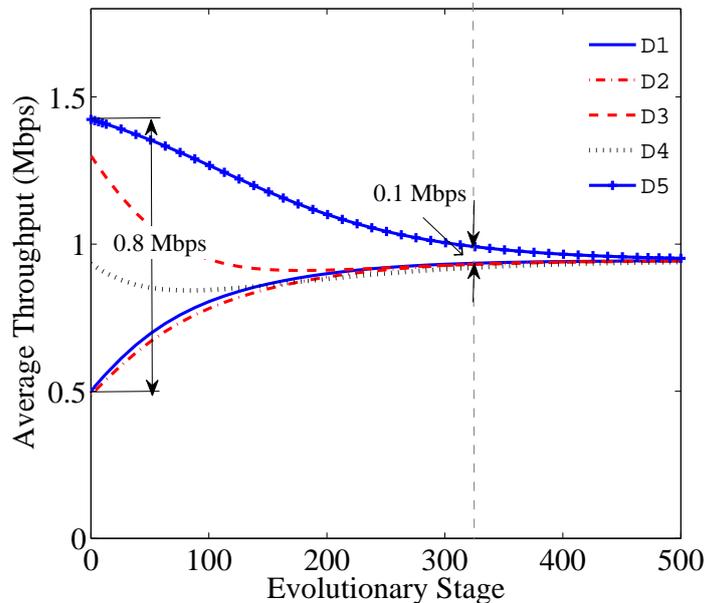


Figure 6: Convergence of the throughput per cell with the growth of evolution stages, where  $N = 150$  and  $M = 5$ .

Fig. 7 shows the stabilization process of the size of each cluster along with evolution over time, corresponding to Fig. 6. We can see the difference of the cluster sizes decreases fast with the evolutionary clustering. However, the cluster sizes do not converge to the same size, as opposed to the throughput curves of the clusters which evolve to converge, as shown in Fig. 6. The reason for this lies in the network geometry of this particular simulation, where the femtocells are geographically unevenly distributed. As a result, the cluster sizes vary so as to balance interference in each cluster.

Another interesting finding in Fig. 7 is that after the cluster sizes stabilize, the throughput of the clusters continues converging. Specifically, by 320 evolution stages, the cluster sizes start to stabilize, while the maximum throughput difference is still about 0.1 Mbps. After that, local adjustment of clustering is happening without significantly changing the cluster sizes. The clustering evolution enables the clusters to locally swap their femtocells (e.g., between clusters 4 and 5 in Fig. 7) and adjust their geometry, until each of the clusters is most widely spanned in space and intra-cluster interference is alleviated.

Figs. 6 and 7 also confirm the effectiveness of our proposed framework, as they evidence that the convergence of throughput among the clusters unnecessarily requires the convergence of the sizes of the clusters. Our proposed framework, especially the cluster conformation design of (1), targets to balance the throughput (or in other words, balance the interference), rather than the cluster sizes.

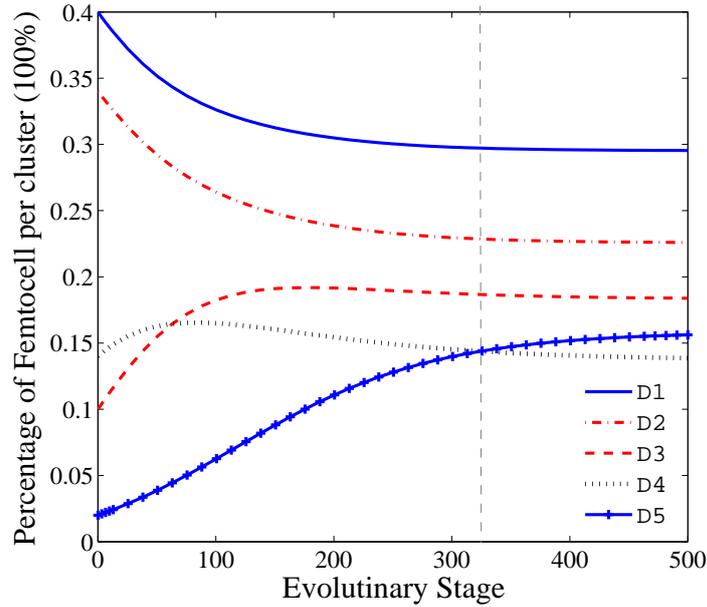


Figure 7: Convergence of the cluster size per cluster with the growth of evolution stages, where  $N = 150$  and  $M = 5$ .

## 5 Prospective standardisations

Given the superior performance, the new hybrid network control is promising for future 5G systems. Standardisations will be important, focusing on the three interfaces of  $I_c$ ,  $I_d$ , and  $I_{c/d}$ . The cycle that the centralised speciation takes place needs to be specified, which has strong impact on the convergence speed of the network evolution, as described in Section 3.1.

Standardisations will also be required on the utility functions described in Section 3.2, which are critical for the convergence of the distributed power control automation, as well as the effectiveness of interference suppression.

Other standardisation efforts will be on identifying typical probability distribution functions for the centralised controller to describe the distribution of data rate in each cluster, as described in Section 3.1. This will be important to leverage the complexity of the centralised SDN controller, as well as the reliability of the network evolution.

## 6 Conclusions

In this article, we proposed to decouple 5G network control into two layers, namely, centralised coordination at the SDN controller and distributed adaptation of base stations. Delay-tolerant topology control and delay-sensitive interference suppression are respectively handled by the two layers. The decoupling was based on a new evolutionary view of 5G network control. The decoupled network control is able to alleviate the requirement of delay and complexity, and therefore enhance network scalability. Case studies show that our decoupled network control is effective for timely interference mitigation and reliable topology management. The stability and scalability of our approach are also demonstrated.

## Acknowledgment

This work was partly supported by National Natural Science Foundation of China under the Grant No.61701034 and No.61531007 and partly supported by China Postdoctoral Science Foundation under the Grant No 2017M620696.

## Bibliography

- [1] Alliance, N. G. M. N. (2012); Small cell backhaul requirements, *white paper*, June, 2012.
- [2] Bemporad, A.; Heemels, M.; Johansson, M. (2010); *Networked control systems*, 406, Berlin: Springer, 2010.
- [3] DOCOMO (2014); DOCOMO 5G White Paper: 5G Radio Access: Requirements, *Concept and Technologies*, 2014.
- [4] Gudipati, A.; Perry, D.; Li, L. E.; Katti, S. (2013); SoftRAN: Software defined radio access network, *Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking*, 25-30, 2013.
- [5] Kim, H.; Feamster, N. (2013); Improving network management with software defined networking. *IEEE Communications Magazine*, 51(2), 114-119, 2013.
- [6] Lin, S.; Tian, H. (2013); Clustering based interference management for QoS guarantees in OFDMA femtocell, *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, 649-654, 2013.
- [7] Ni, W.; Collings, I. B. (2012); Indoor wireless networks of the future: adaptive network architecture, *IEEE Communications Magazine*, 50(3), 130-137, 2012.
- [8] Ni, W.; Collings, I. B. (2013). A new adaptive small-cell architecture, *IEEE Journal on Selected Areas in Communications*, 31(5), 829-839, 2013.
- [9] Ni, W.; Collings, I. B.; Lipman, J.; Wang, X.; Tao, M.; Abolhasan, M. (2015); Graph theory and its applications to future network planning: Software-defined online small cell management, *IEEE Wireless Communications*, 22(1), 52-60, 2015.
- [10] Ni, W.; Liu, R.P. ; Collings, I.B.; Wang, X. (2013), Indoor cooperative small cells over Ethernet, *Communication Magazine, IEEE*, 51(9), 100-107, 2013.
- [11] Pentikousis, K.; Wang, Y.; Hu, W. (2013); Mobileflow: Toward software-defined mobile networks, *IEEE Communications magazine*, 51(7), 44-53, 2013.
- [12] Shapley, L. S. (1971); Cores of convex games, *International journal of game theory*, 1(1), 11-26, 1971.
- [13] Wen, T.; Zhu, P. (2013); 5G: A technology vision, *Huawei*, 2013.

# New Failure Mode and Effects Analysis based on D Numbers Downscaling Method

B. Liu, Y. Hu, Y. Deng

## Baoyu Liu

College of Information Science and Technology  
Jinan University, Guangzhou, China  
Baoyu.L@hotmail.com

## Yong Hu

Big Data Decision Institute,  
Jinan University, Tianhe, Guangzhou, 510632, China  
henryhu200211@163.com

## Yong Deng\*

1. Big Data Decision Institute,  
Jinan University, Tianhe, Guangzhou 510632, China  
2. Institute of Fundamental and Frontier Science,  
University of Electronic Science and Technology of China  
Chengdu, 610054, China  
\*Corresponding author: prof.deng@hotmail.com

**Abstract:** Failure mode and effects analysis (FMEA) is extensively applied to process potential faults in systems, designs, and products. Nevertheless, traditional FMEA, classical risk priority number (RPN), acquired by multiplying the ratings of occurrence, detection, and severity, risk assessment, is not effective to process the uncertainty in FMEA. Many methods have been proposed to solve the issue but deficiencies exist, such as huge computing quality and the mutual exclusivity of propositions. In fact, because of the subjectivity of experts, the boundary of two adjacent evaluation ratings is fuzzy so that the propositions are not mutually exclusive. To address the issues, in this paper, a new method to evaluate risk in FMEA based on D numbers and evidential downscaling method, named as D numbers downscaling method, is proposed. In the proposed method, D numbers based on the data are constructed to process uncertain information and aggregate the assessments of risk factors, for they permit propositions to be not exclusive mutually. Evidential downscaling method decreases the number of ratings from 10 to 3, and the frame of discernment from  $2^{10}$  to  $2^3$ , which greatly reduce the computational complexity. Besides, a numerical example is illustrated to validate the high efficiency and feasibility of the proposed method.

**Keywords:** Failure mode and effects analysis; Dempster-Shafer evidence theory; basic belief assignment; belief function; risk priority number; D numbers; evidential downscaling; aggregate assessment

## 1 Introduction

Failure mode and effects analysis (FMEA) was first developed to assess known and potential faults and prevent them from happening in the 1960s [5]. It is an efficient and powerful technology in engineering and management fields, including defining, identifying, and removing known and potential faults, errors, and risk from the system, process, design as well as service. [1, 51]. Besides, in other fields, such as risk management, healthcare management and engineer design, FMEA plays an important role [2, 8, 10, 11, 15, 24, 36]. Furthermore, FMEA not only identifies factors that induce fault but also classifies the likelihood and severity of such fault [11].

A traditional method in FMEA is risk priority number (RPN), which is acquired by multiplying the grades of occurrence, severity, and detection. Thus, how to aggregate the assessments of three risk factors is a key issue, especially when the evaluated information given by experts is uncertain. These years, a large number of approaches have been proposed to improve FMEA methodology [5, 11, 23, 26, 37, 43]. It's inevitable to deal with uncertainty in real world [20–22, 29, 44]. Due to the efficiency to handle linguistic information of fuzzy sets [13, 18], an intuitionistic fuzzy approach for FMEA is proposed in [15], which offers some advantages over earlier models as it accounts for degrees of uncertainty in relationships among various criteria or options, specifically when relations cannot be expressed in definite numbers. In addition, an integrating hesitant 2-tuple linguistic term sets and an extended QUALIFLEX approach is proposed by Liu H C et al. in [13]. Furthermore, as is known to all, D numbers [9] and grey theory are two popular methods to process the uncertain information. Therefore, based on the two theories, plenty of methods are presented [23].

Classical RPN is criticized due to its several shortcomings. For example, the importance weights of three risk factors are taken into consideration. Besides, classical RPN can do nothing about the uncertain information, etc. Aiming at these problems, a method which is effective to dispose the uncertainty of assessment is proposed by Chin et al. [4]. However, the algorithm is too complex computationally, the reason is that the numeric ratings of every risk factor are from 1 to 10 so that the number of frame of discernment is  $2^{10}$ , which greatly increase the computational load [11]. In other to reduce the computational complexity, an evidential downscaling method is proposed in [11]. Nevertheless, the D-S combination theory it uses requests that propositions are exclusive mutually. As a matter of fact, traditional FMEA ratings are obtained by subjective judgment of the experts. Therefore, the boundary of two adjacent ratings is fuzzy, hence the propositions are not actually mutually exclusive. For purpose of solving this problem, in this paper, a new method to evaluate risk in FMEA based on D numbers and evidential downscaling method, named as D numbers downscaling method, is proposed. In the proposed method, on the one hand, the evidential downscaling is utilized to decrease the frame of discernment so that greatly reduces the computational complexity. On the other hand, according to the data, D number is constructed to processing uncertain information and aggregate the assessment of risk factors because D numbers permit propositions that are not exclusive mutually.

The rest of the paper is organized as follows. Key concepts and previous theories are briefly reviewed in Section 2. A new method to evaluate risk based on D numbers downscaling method is proposed in Section 3. A numerical example is illustrated to show the feasible of the proposed method in Section 4. A brief conclusion is drawn in Section 5.

## 2 Preliminaries

### 2.1 Risk priority number

The real systems are very complicated to be modelled [14,27,46]. Risk priority number (RPN) is a traditional and typical method to model and evaluate risk in FMEA of the complicated systems, which is acquired by multiplying the grades of occurrence assessment (O), severity assessment (S), and detection assessment (D). That is

$$RPN = OSD,$$

where O stands for the probability of occurrence of failure mode, S refers to the severity of failure mode and D refers to the probability of failure being detected. The three risk factors are evaluated by FMEA experts using a 1 to 10 numeric scale, as expressed in Tables 1 to 3 [11].

The larger RPN is, the more important degree it is supposed to be assigned, referring to the failure mode should be more priority to be corrected.

However, as a traditional method for FMEA, classical RPN has been controversial all the time for many reasons, and the most important ones are listed as follows:

- The three risk factors have the same importance in RPN, but in practical applications of FMEA, they might be treated with different weight.
- Other factors are not taken into consideration, such as economy factor.
- Different combination of risk factors might obtain the same RPN, nevertheless, the potential risk might be totally different.
- The mathematical formula for calculating RPN lacks the scientific basis. There is not any proof to clarify the reason why O, S and D should be multiplied to obtain RPN.
- In fact, the scores of the three factors are difficult to be determined accurately. Therefore, FMEA experts usually provide different types of assessment information, some of which are uncertain and incomplete data.

### 2.2 Dempster-Shafer evidence theory

Dempster-Shafer evidence theory (D-S evidence theory) constructs a basic probability assignment(BPA) in the frame of discernment. Through combining BPAs the imprecise and uncertain information can be fused [5]. With its rapid development, it is regarded as an important method that is extensively applied in many fields such as complex networks and systems [8, 10, 12, 19, 45, 50, 52], multisource information fusion [6, 7, 17, 28, 41, 42, 49], uncertainty modelling [1, 2, 16, 48, 53], pattern recognition [30, 31] and Imprecise payoff [34, 35].

**Definition 1.** Let  $\Theta = \{H_1, H_2, \dots, H_N\}$  be a finite nonempty set, which is composed of N mutually exclusive and exhaustive elements. Denote  $P(\Theta)$  as the power set composed of  $2^N$  elements of  $\Theta$ . The BBAs function is defined as a mapping of the power set  $P(\Theta)$  to a number between 0 and 1, that is,  $m : P(\Theta) \rightarrow [0, 1]$ , and which satisfies the following conditions:

$$m(\emptyset) = 0; \tag{1}$$

$$\sum_{A \subseteq P(\Theta)} m(A) = 1. \tag{2}$$

The mass  $m(A)$  represents how strongly the evidence supports A.

**Definition 2.** Two bodies of evidence A and B of  $\Theta$  can be used to calculate the belief level for some new hypothesis C shown as follows: The measure of conflict K that is also called as the conflict coefficient between A and B is given as:

$$K = \sum_{A \cap B = \emptyset, \forall A, B \subseteq \Theta} m_i(A)m_{i'}(B); \tag{3}$$

and the mass function after combination is:

$$m(C) = \begin{cases} \frac{\sum_{A \cap B = C, \forall A, B \subseteq \Theta} m_i(A)m_{i'}(B)}{1 - K}, & \text{when } A \neq \emptyset \\ 0, & \text{when } A = \emptyset, \end{cases} \tag{4}$$

where  $K$  reflects the conflict between two bodies of evidence. Absolutely,  $0 \leq K \leq 1$ .  $K = 0$  shows the absence of conflict between two bodies of evidence.  $K = 1$  shows complete conflict between  $m_i$  and  $m_{i'}$ . When  $K = 1$ , the Dempster's rule of combination is no longer applicable.

### 2.3 The evidential downscaling method

An evidential downscaling method is proposed in [11], which is based on Euclidean distance from multi-scale to construct the BBAs. The new method, decreasing the frame of discernment from  $2^{10}$  to  $2^3$ , is shown as follows:

Step 1. Calculate the group assessment into a crisp number with weighted average method.

Step 2. Construct the BBAs by Definition 2 with the Euclidean distances between group assessment and multi-scale ratings. When calculating the distance, an evidential downscaling method is presented to divide the rating into three scales: '10' for 'bad', '1' for 'good', and '5' for 'bad or good' because rating 10 expresses the most serious degree, rating 1 expresses that failure is almost impossible and rating 5 expresses the moderate degree, which decreases the number of frame of discernment from  $2^{10}$  to  $2^3$ .

The BBAs can be constructed as follows:

$$m(\{B\}) = \frac{d(G)}{d(G) + d(B, G) + d(G)}, \quad (5)$$

$$m(\{G\}) = \frac{d(B)}{(d(G) + d(B, G) + d(G))}, \quad (6)$$

$$m(\{B, G\}) = 1 - m(\{B\}) - m(\{G\}). \quad (7)$$

Step 3. Combine the BBAs of risk factors by Dempster's combination rule.

Step 4. Acquire the risk priority according to the aggregation of 'bad'. Let  $m(\{B\})_{OSD}$ ,  $m(\{G\})_{OSD}$  and  $m(\{B, G\})_{OSD}$  be the aggregation of three risk factors for 'bad', 'good', and 'bad or good'. The final assessment of 'bad' and 'good' can be calculated as follows:

$$m(\{B\})'_{OSD} = m(\{B\})_{OSD} + \frac{m(\{B, G\})_{OSD}}{2}, \quad (8)$$

$$m(\{G\})'_{OSD} = m(\{G\})_{OSD} + \frac{m(\{B, G\})_{OSD}}{2}. \quad (9)$$

### 2.4 D numbers [9]

D numbers is an effective method to deal with uncertain information, which overcomes the shortcomings of Dempster-Shafer (D-S) theory. Nowadays, D number is extensively used in many fields such as dependence assessment, emergency management, and aggregating operator [33, 54, 55]. It can be defined as follows:

**Definition 3.** Let  $\Omega$  be a finite nonempty set, D number is a mapping  $\mathbf{D} : \Omega \rightarrow [0, 1]$ , such that

$$\sum_{B \subseteq \Omega} D(B) \leq 1 \text{ and } D(\emptyset) = 0, \quad (10)$$

where  $\emptyset$  is an empty set and  $B$  is a subset of  $\Omega$ . More importantly, different from the concept of frame of discernment in D-S theory, the elements of set  $\Omega$  do not require to be mutually exclusive in D numbers. If  $\sum_{B \subseteq \Omega} D(B) = 1$ , the information is considered to be complete; otherwise, the information is regarded to be incomplete.

For a set  $\Omega = \{b_1, b_2, \dots, b_i, \dots, b_n\}$ , where  $b_i \neq b_j$ , if  $i \neq j$ . Then, a special D number can be expressed by

$$D(\{b_1\}) = v_1,$$

$$D(\{b_2\}) = v_2,$$

..... ,

$$D(\{b_i\}) = v_i,$$

..... ,

$$D(\{b_n\}) = v_n$$

or  $D = \{(b_1, v_1), (b_2, v_2), \dots, (b_i, v_i), \dots, (b_n, v_n)\}$ , where  $v_i > 1$  and  $\sum_{i=1}^n v_i \leq 1$ . For the detailed information about D numbers combination rule, please refer [9]. A combination rule, a kind of add operation, is proposed to combine two D numbers.

### 3 The proposed method

As a traditional method for FMEA, classical RPN is acquired by multiplying the rating of three risk factors. Nevertheless, classical RPN is criticized for several disadvantages mentioned in 2.1. With the aim of overcoming these shortcomings, a large number of methods to evaluate risk for FMEA are proposed. However, existing methods either have a huge computing quantity, such as the Chinś method, or require the propositions to be mutually exclusive, like the method proposed by Du Y et al.. As a matter of fact, traditional ratings are divided by the subjective judgments of the experts which are based on the individual experience. Therefore, the boundary of two adjacent ratings is fuzzy, which means that two propositions in FMEA are not mutually exclusive. Hence, it is obviously unreasonable that Dempster-Shafer combination theory is used to aggregate assessment in the evidential downscaling method proposed in [11]. Aiming at these problems, a new method to evaluate risk in FMEA based on D numbers and evidential downscaling method is proposed. Based on the three risk factors as well, the proposed method, for one thing, decreases the number of the frame of discernment from  $2^{10}$  to  $2^3$  by making use of the evidential downscaling method, which greatly reduce the computational complexity. For another, D numbers are utilized to manage the uncertain information. Because D numbers allow the propositions to not be exclusive, it is reasonable that D numbers are used to process uncertain information and aggregate the assessments. Furthermore, in consideration of the fuzzy information given by the experts on different failure modes, weighted averages are calculated to substitute them in the proposed method. Suppose there are  $N$  failure modes and  $M$  experts, the specific steps are shown as follows: The failures with higher values of ‘bad’ are assumed to be more important and should be given higher priorities.

- Step 1. Calculate the mathematical expectation of the score given by experts for evaluating each risk factor by Eq. (11).

**Definition 4.** Let  $S_l^{i,j}$  be the mathematical expectation of the score  $S$  of failure mode  $i$  given by expert  $j$  in the assessment of risk factor  $l$ , it is calculated as follows:

$$S_l^{i,j} = \begin{cases} S, & \text{if } S \text{ is integer} \\ a \times c\% + b \times d\%, & \text{if } S \text{ is } (a : c\% , b : d\%) \text{ with } c\% + d\% = 100\% \\ a \times c\% + \frac{1}{9} \times (55 - a) \times (1 - c\%), & \text{if } S \text{ is } (a : c\%) \text{ but } 0 < c\% < 100\% \\ 5.5, & \text{if } S \text{ is missing} \end{cases} \quad (11)$$

where  $i = 1, \dots, N$ ,  $j = 1, \dots, M$ ,  $l = O, S, D$ .

Step 2. Calculate the Euclidean distances between group assessment and multi-scale ratings. Consistent with the evidential downscaling method mention in 2.3, then the new score  $SG_l^{i,j}$  of each factor of each failure mode after downscaling can be obtained as follows:

$$d(B)_l^{i,j} = |S_l^{i,j} - 10|, \quad (12)$$

$$d(G)_l^{i,j} = |S_l^{i,j} - 1|, \quad (13)$$

$$d(B, G)_l^{i,j} = |S_l^{i,j} - 5|, \quad (14)$$

$$SG_l^{i,j} = \frac{d(B)_l^{i,j}}{(d(G)_l^{i,j} + d(B, G)_l^{i,j} + d(G)_l^{i,j})}, \quad (15)$$

where  $i = 1, \dots, N$ ,  $j = 1, \dots, M$ ,  $l = O, S, D$ .

Step 3. Construct D numbers on the basis of  $SG_l^{i,j}$  in the light of Definition 5.

**Definition 5.** Let  $D_{F_i}^{E_j}$  be the D number of failure mode  $i$  of experts  $j$ , which stands for the three assessments of experts  $j$  towards failure mode  $i$ , is modeled as follows:

$$D_{F_i}^{E_j} = \{(b_O^{ij}, v_O), (b_S^{ij}, v_S), (b_D^{ij}, v_D)\} \quad (16)$$

with

$$b_l^{i,j} = \lambda_j \times SG_l^{i,j} \quad (17)$$

where  $\lambda_j$  is the weight for expert  $j$ , and  $v_l$  ( $l = O, S, D$ ) is the weight of the risk factor standing for the importance of it in FMEA. In addition,  $i = 1, \dots, N$ ,  $j = 1, \dots, M$ ,  $l = O, S, D$ .

Step 4. According to the processes of  $D_{F_i}$  numbers mentioned in [9] and Definition 6, for each failure mode, the combination of D numbers can be calculated and the  $I(D)$  of each failure mode can be obtained as well. Most importantly, the failure modes with the lower values of  $I(D)$  are assumed to be more important and should be given higher priorities.

**Definition 6.** Let  $D_{F_i}$  be the D number of failure mode  $i$  ( $i = 1, \dots, N$ ) aggregate the assessment of  $M$  experts, thus, it can be calculate as follows:

$$D_{F_i} = D_{F_i}^{E_1} \oplus D_{F_i}^{E_2} \oplus \dots \oplus D_{F_i}^{E_j} \oplus \dots \oplus D_{F_i}^{E_M} \quad (18)$$

Figure 1 is the sub-flowsheet of the proposed method, that is, the procedure of downscaling and constructing D number. Figure 2 is the general flow-chart.

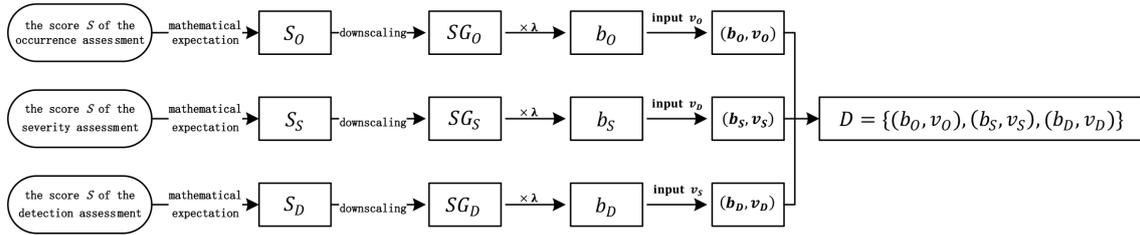


Figure 1: The procedure of downscaling and constructing D number

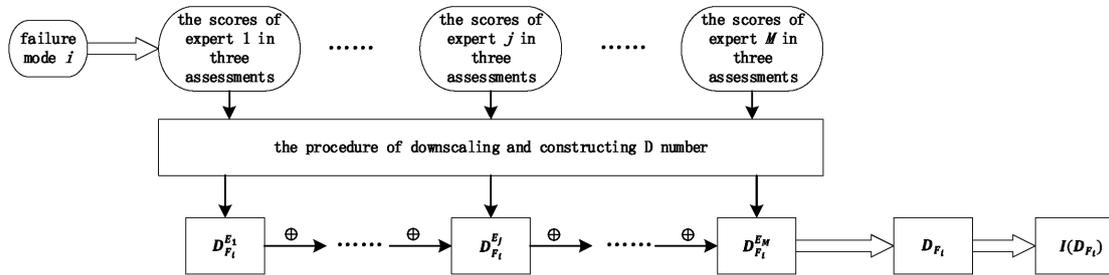


Figure 2: The general flow-chart of the proposed method

Table 1: Occurrence assessment by FMEA team members

Failure mode	Expert1	Expert2	Expert3	Expert4	Expert5
1	1	1		1	1
2	1 : 50%, 2 : 50%	1	1	1	1
3	2	2 : 90%	2	2	2
4	8	8	8 : 80%, 9 : 20%	8	8
5	6	6	6	6	6
6	2	13	2	2	2-3
7	2	2	2	9	2
8	1	1 : 75%, 2 : 25%	1	1	1
9	3	3	3	3	3
10	1 : 80%, 2 : 20%	1	1	1	1 - 2 : 85%, 3 : 15%
11	4	4	4	3 - 4 : 75%, 5 : 25%	4
12	9	9	9	9	9
13	8	8 : 80%	8	8	8
14	3	3	4	3	3
15	3	3	3	3	3 : 70%, 4 : 30%
16	1	1	1	1	1
17	3 - 5 : 90%, 6 : 10%	4	4	4	4
18	2	2	2 : 90%	2	2
19	7	7	7	7	7:80%
20	9	9	9	7 : 30%, 8 - 9 : 70%	9
21	9	8 - 9	9	9	9

Table 2: Severity assessment by FMEA team members

Failure mode	Expert1	Expert2	Expert3	Expert4	Expert5
1	7 : 20%, 8 : 80%	8	8	6-7 : 50%, 8-9 : 50%	8
2	8	8	8	8	8
3	7-9 : 90%	8	6-8	8	8
4	8	8	8	8	7-9 : 80%
5	8	7-9 : 90%	8	8	8
6	8	8	8	6-8	8
7	9 : 75%, 8 : 25%	9	9	9	9
8	4	4	4	4 : 50%, 5 : 50%	3-5 : 75%, 6-7 : 25%
9	2	2	2	2	2
10	2	2	1-2 : 60%, 3-4 : 60%	2	2
11	2	2-3	2	2	2
12	3	3	3 : 60%, 4 : 40%	3	3
13	2-3 : 80%, 3-4 : 20%	3	3	3	3
14	7	8	7	7	7
15	3	3	3	3	3
16		8	8	8	8
17	8	4	8	8	8
18	7	7	7	7	7
19	2	1-2 : 75%, 2-3 : 25%	2	2	2
20	8			8	8
21	3	3	3	3	3

## 4 Numerical example

Suppose that there are 5 experts evaluating risk in FMEA, who give their assessments on the three risk factors of 21 failure modes as shown in Table 1 to 3. The weights for the five experts are assumed to be 0.3, 0.3, 0.2, 0.1 and 0.1.

Take failure mode 1 of expert 1 for example. As shown in Table 1, three ratings the expert 1 gives are 1, (7:20%, 8:80%), and 3. Thus, according to Eq. (11), the mathematical expectation of score are 1, 7.8, and 3.

Then, the Euclidean distances between it and ‘bad’, ‘good’ and ‘bad or good’ are calculating as in Table 4.

Therefore, the new scores can be obtained as follows:

$$SG_O^{11} = \frac{d(B)_O^{11}}{d(G)_O^{11} + d(B, G)_O^{11} + d(G)_O^{11}} = \frac{9}{9 + 0 + 4} \approx 0.6923.$$

Table 3: Detection assessment by FMEA team members

Failure mode	Expert1	Expert2	Expert3	Expert4	Expert5
1	3	3	3	3	3 : 90%
2	3	3	3	3	3
3	4	4	4	4	3-4 : 80%, 5-6 : 20%
4	5	5	5	5	5
5	6	6	6 : 85%, 7 : 15%	6	6
6	1	1 : 85%, 2 : 15%	2		1
7	3	2	2	1-2 : 75%, 3-4 : 25%	2
8	3	3	3 : 80%, 4 : 20%	3	3
9	3	3	3-4 : 60%, 5 : 40%	3	3
10	4	4	4	4	4
11	3 : 70%, 5 : 30%	3	3	3	
12	7	7	7	7	7
13	6	6	6	5-7	6
14	4	4	4	4	4
15	4	4 : 95%	4	4	4
16	3	3	3	3	3
17	5	5	5	5	5
18	7	6-8	7	7	7
19	4	4	4	8-9 : 90%	4
20	4 : 60%	9	9	9	9
21	6	6	4-6	6	4 : 25%, 5-7 : 75%

Table 4: Euclidean distances between it and "bad", "good" and "bad or good"

Bad	Good	Bad or Good
$d(B)_O^{11} =  1 - 10  = 9,$	$d(G)_O^{11} =  1 - 1  = 0,$	$d(B, G)_O^{11} =  1 - 5  = 4;$
$d(B)_S^{11} =  7.8 - 10  = 2.2,$	$d(G)_S^{11} =  7.8 - 1  = 6.8,$	$d(B, G)_S^{11} =  7.8 - 5  = 2.8;$
$d(B)_D^{11} =  3 - 10  = 7,$	$d(G)_D^{11} =  3 - 1  = 2,$	$d(B, G)_D^{11} =  3 - 5  = 2;$

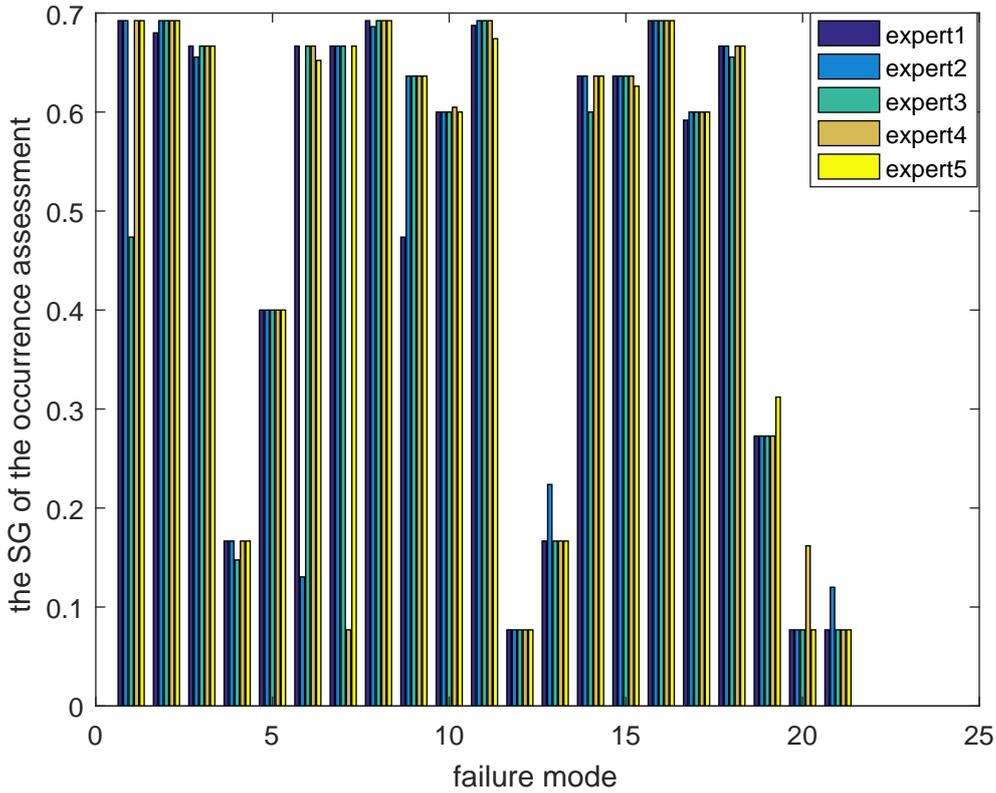


Figure 3: The new score of Occurrence assessment

$$SG_S^{11} = \frac{d(B)_S^{11}}{d(G)_S^{11} + d(B, G)_S^{11} + d(G)_S^{11}} = \frac{2.2}{2.2 + 6.8 + 2.8} \approx 0.1864.$$

$$SG_D^{11} = \frac{d(B)_D^{11}}{d(G)_D^{11} + d(B, G)_D^{11} + d(G)_D^{11}} = \frac{7}{7 + 2 + 2} \approx 0.6364.$$

The data of the other three assessments is treated in the same way, among which the result of occurrence assessment with the new scores are shown in Figure 3. As shown in Figure 3, the higher initial rating tends to the lower new score while the lower rating tends to the higher one.

Then, the first element  $b$  can be calculated by multiply the weight of the expert 1 and the new score:

$$b_O^{11} = \lambda_1 \times SG_O^{11} = 0.3 \times 0.6923 = 0.2077.$$

Table 5: The representation of D numbers for failure mode 1 ( $F_1$ )

$F_1$	D numbers
Expert 1	$D_{F_1}^{E_1} = \{(0.2077, \frac{1}{3}), (0.0559, \frac{1}{3}), (0.1909, \frac{1}{3})\}$
Expert 2	$D_{F_1}^{E_2} = \{(0.2077, \frac{1}{3}), (0.0500, \frac{1}{3}), (0.1909, \frac{1}{3})\}$
Expert 3	$D_{F_1}^{E_3} = \{(0.0947, \frac{1}{3}), (0.0333, \frac{1}{3}), (0.1273, \frac{1}{3})\}$
Expert 4	$D_{F_1}^{E_4} = \{(0.0692, \frac{1}{3}), (0.0217, \frac{1}{3}), (0.0636, \frac{1}{3})\}$
Expert 5	$D_{F_1}^{E_5} = \{(0.0692, \frac{1}{3}), (0.0167, \frac{1}{3}), (0.0627, \frac{1}{3})\}$

$$b_S^{1,1} = \lambda_1 \times SG_S^{1,1} = 0.3 \times 0.6923 = 0.0559.$$

$$b_D^{1,1} = \lambda_1 \times SG_D^{1,1} = 0.3 \times 0.6923 = 0.1909.$$

With the same process, the  $b$  of other D numbers can be obtained. Next, the data is treated using the D numbers. It needs to be emphasized that the weight of each rich factor is treated as the second parameter  $v$ . It is worth mentioning that the judgments for the relative importance of each criterion are various from different expert.

In order to compare with other methods which do not take it into consideration and prove the feasibility of the proposed method as well, the importance of three factors are assumed to be same. That is,  $v$  is supposed to be identically equal to  $\frac{1}{3}$ .

Therefore, the assessment result being disposed through above process are expressed in the forms of D numbers. According to the data of expert 1, a D number  $D_{F_1}^{E_1}$  is constructed, where  $D_{F_1}^{E_1} = \{(0.2077, \frac{1}{3}), (0.0559, \frac{1}{3}), (0.1909, \frac{1}{3})\}$ . Similarity to  $D_{F_1}^{E_1}$ , According to the data of expert 2, expert 3, expert 4 and expert 5, four D numbers  $D_{F_1}^{E_2}$ ,  $D_{F_1}^{E_3}$ ,  $D_{F_1}^{E_4}$  and  $D_{F_1}^{E_5}$  are generated. Table 4 shows these D numbers. In the same method, for failure mode  $F_i$ , and expert  $E_j$ , each D number  $D_{F_i}^{E_j}$  can be constructed, too.

For each failure mode, all experts' data represented by D numbers are combined according to the combination rule of D numbers mentioned in [9]. Take the failure mode 1 ( $F_1$ ) for example, the integrated assessment of three experts is the aggregation of  $D_{F_1}^{E_1}$ ,  $D_{F_1}^{E_2}$ ,  $D_{F_1}^{E_3}$ ,  $D_{F_1}^{E_4}$  and  $D_{F_1}^{E_5}$ ,

$$D_{F_1} = D_{F_1}^{E_1} \oplus D_{F_1}^{E_2} \oplus D_{F_1}^{E_3} \oplus D_{F_1}^{E_4} \oplus D_{F_1}^{E_5}.$$

In accordance with the processes of  $D_{F_i}$  numbers mentioned in [9],  $D_{F_1}$  is calculated. Therefore, the  $I(D_{F_1})$  of  $D_{F_1}$  can be obtained. Using the same method,  $D_{F_i}$  and  $I(D_{F_i})$  are acquired, which are shown in Table 5. In the meantime, according to  $I(D_{F_i})(i = 1, 2, \dots, 21)$ , the ranking of 21 failure modes are obtained, where the failure modes with the lower values of  $I(D)$  are given higher priorities.

In Table 6, the risk priority rankings of the evidential downscaling method are shown. Compared with the ranking of the evidential downscaling method, the proposed method's is similar. As seen from Table 6, failure mode 3, failure mode 9, failure mode 12, failure mode 13, failure mode 15, failure mode 18, failure mode 20 and failure mode 21 have the same risk priority ranking in both methods. Besides, it is indicated that the five of highest risk priority rankings are

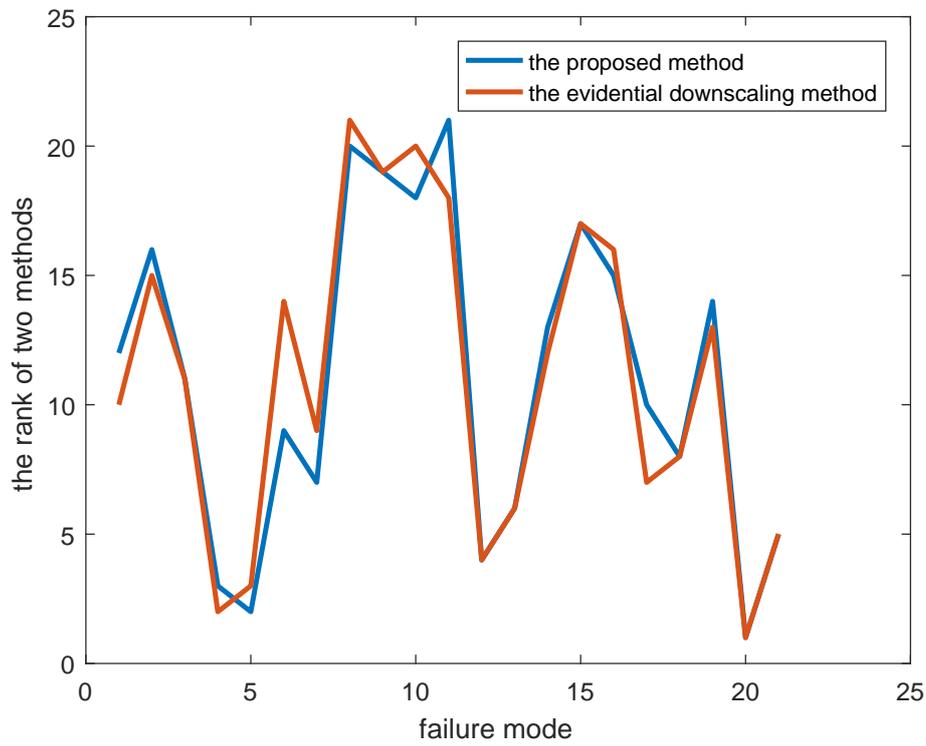


Figure 4: The comparison of risk priority ranking by two method

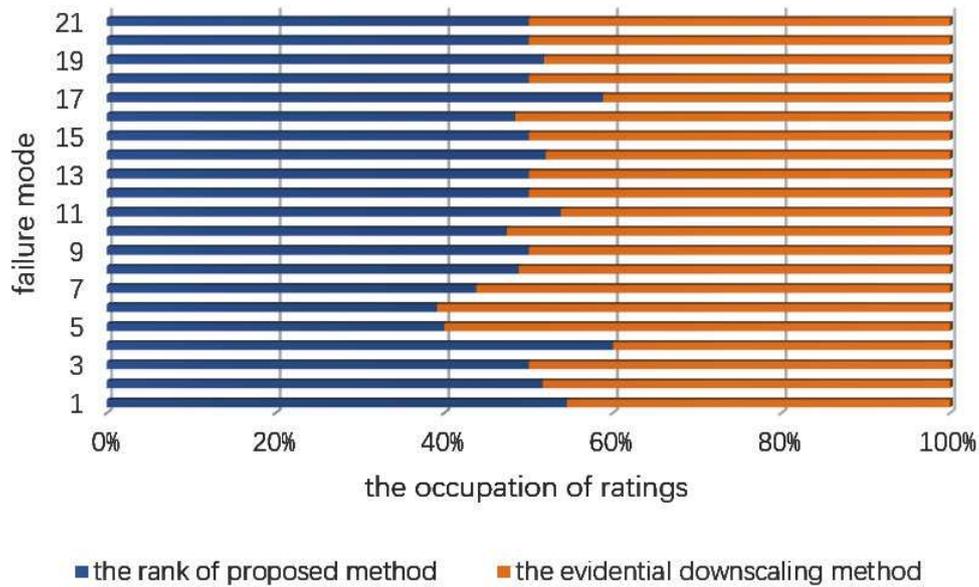


Figure 5: The occupation of ratings between two method

failure mode 20, 4, 5, 12 and 21, which means that these 5 faults are most likely to occur in both two methods. Moreover, in both two methods, the five of the lowest priorities are failure mode 8, 9, 10, 11 and 15, indicating that these 5 failures are supposed to be.

Figure 4 shows the comparison of risk priority rankings by the proposed method and the

Table 6: The comparison of risk priority ranking by two methods

Failure mode	Result	The rank of proposed method	The evidential downscaling method
1	0.06634	12	10
2	0.07464	16	15
3	0.06597	11	11
4	0.04524	3	2
5	0.04210	2	3
6	0.06309	9	14
7	0.05789	7	9
8	0.08779	20	21
9	0.08743	19	19
10	0.08616	18	20
11	0.08824	21	18
12	0.04596	4	4
13	0.05568	6	6
14	0.06670	13	12
15	0.08518	17	17
16	0.06896	15	16
17	0.0633	10	7
18	0.06064	8	8
19	0.06781	14	13
20	0.02350	1	1
21	0.05345	5	5

evidential downscaling method, in which the X-axis shows the rankings and Y-axis shows the failure modes. As is shown in Figure 4, it indicates that the rank result generated by proposed method has similar trend with the evidential downscaling method, which proves the proposed method is valid. However, in consideration that it is out of reality to request the propositions to be exclusive mutually, the proposed method is more reasonable. In 5, the occupation of ratings between two methods for each failure mode is shown. From Figure 5, the edge of the area around with different color is almost near the midline of the graph, which proves that the proposed method similarly has a good effect for risk assessment.

## 5 Conclusion

The conventional FMEA method has been criticized for its deficiencies especially in the evaluation of risks of failure mods and computation of classical RPN. A large number of new methods for FMEA have been proposed. Nevertheless, some deficiencies still exist in these methods, such as huge computing quality. Although the evidential downscaling method presented by Du Y et al. is capable to greatly reduce the amount of calculation, it is not reasonable that the propositions are requested to be exclusive mutually. Therefore, a new method to evaluate risk in FMEA based on D numbers downscaling method is proposed with the purpose of solving this problem.

Obviously, the proposed method can handle the uncertain information well in FMEA. Compared with the traditional RPN, the proposed method not only can dispose the scores of three risk factors even if they are uncertain, but also takes into consideration the relative importance weights of risk factors. Therefore, the result of the proposed method is more reliable. In comparison to the existing method, the proposed has more succinct calculation formulas, which has a smaller amount of calculation so that it is more worthy of promotion. Furthermore, because of

the use of D numbers, the propositions are not required to be exclusive mutually, which proves the proposed method is more reasonable than the evidential downscaling method. In conclusion, the proposed method is a feasible and efficient method to risk assessment in failure mode and effects analysis. In addition, although the proposed method takes the weights of three factors into consideration, it does not make sure the method to obtain the weights, which would be the further exploration.

## Acknowledgment

The authors greatly appreciate the reviewers' suggestions and the editor's encouragement. The work is partially supported by National Natural Science Foundation of China (Grant Nos. 61573290, 61503237).

## Bibliography

- [1] Abellán, J., Mantas, C. J., Castellano, J. G., (2017); A random forest approach using imprecise probabilities, *Knowledge-Based Systems*, 134, 72–84, 2017.
- [2] Abellán, J., (2017); Analyzing properties of deng entropy in the theory of evidence, *Chaos Solitons & Fractals*, 95, 195–199, 2017.
- [3] Bowles, J. B., Pelaez, C. E., (1995); Fuzzy logic prioritization of failures in a system failure mode, effects and criticality analysis, *Reliability Engineering System Safety*, 50(2), 203–213, 1995.
- [4] Chin, K. S., Wang, Y. M., Gary, K. K. P., Yang, J. B., (2009); Failure mode and effects analysis using a group-based evidential reasoning approach, *Computers & Operations Research*, 36 (6), 1768–1779, 2009.
- [5] Dempster, A. P., (1967). Upper and lower probabilities induced by a multivalued mapping, *Annals of Mathematical Statistics*, 38 (2), 325–339, 1967.
- [6] Deng, X., Han, D., Dezert, J., Deng, Y., Shyr, Y., (2016); Evidence Combination From an Evolutionary Game Theory Perspective, *IEEE Transactions on Cybernetics*, 46(9), 2070–2082, 2016.
- [7] Deng, X., Liu, Q., Deng, Y., (2016); Matrix games with payoffs of belief structures, *Applied Mathematics and Computation*, 273, 868–879, 2016.
- [8] Deng, X., Liu, Q., Deng, Y., Mahadevan, S., (2016); An improved method to construct basic probability assignment based on the confusion matrix for classification problem, *Information Sciences*, 340, 250–261, 2016.
- [9] Deng, Y., (2012). D numbers: theory and applications, *Journal of Information and Computational Science*, 9 (9), 2421–2428, 2012.
- [10] Dong, Y., Wang, J., Chen, F., Hu, Y., Deng, Y., (2017); Location of facility based on simulated annealing and ZKW algorithms, *Mathematical Problems in Engineering*, Article ID 4628501, 2017.
- [11] Du, Y., Lu, X., Su, X., Hu, Y., Deng, Y., (2016); New failure mode and effects analysis: An evidential downscaling method, *Quality & Reliability Engineering*, 32 (2), 737–746, 2016.

- 
- [12] Fei, L.; Wang, H.; Chen, L.; Deng, Y. (2017); A new vector valued similarity measure for intuitionistic fuzzy sets based on OWA operators, *Iranian Journal of Fuzzy Systems*, accepted, 2017.
- [13] Goyal, R. K., Kaushal, S., (2016); A constrained non-linear optimization model for fuzzy pairwise comparison matrices using teaching learning based optimization, *Applied Intelligence*, 1–10, 2016.
- [14] Hu, Y., Du, F., Zhang, H. L., (2016); Investigation of unsteady aerodynamics effects in cycloidal rotor using RANS solver, *Aeronautical Journal*, 120(1228), 956–970, 2016.
- [15] Jain, K., (2017). Use of failure mode effect analysis (fmea) to improve medication management process, *International Journal of Health Care Quality Assurance*, 30 (2), 175, 2017.
- [16] Jiang, W., Wang, S., (2017); An uncertainty measure for interval-valued evidences, *International Journal of Computers Communications & Control*, 12 (5), 631–644, 2017.
- [17] Jiang, W., Wang, S., Liu, X., Zheng, H., Wei, B., (2017); Evidence conflict measure based on OWA operator in open world, *PloS one*, 12 (5), e0177828, 2017.
- [18] Jiang, W., Xie, C., Zhuang, M., Tang, Y., (2017); Failure mode and effects analysis based on a novel fuzzy evidential method, *Applied Soft Computing*, 57, 672–683, 2017.
- [19] Kang, B., Chhipi-Shrestha, G., Deng, Y., Mori, J., Hewage, K., Sadiq, R. (2017); Development of a predictive model for clostridium difficile infection incidence in hospitals using gaussian mixture model and dempster-shafer theory, *Stochastic Environmental Research and Risk Assessment*, accepted, 2017.
- [20] Li, C., Mahadevan, S., (2016); An efficient modularized sample-based method to estimate the first-order sobol index, *Reliability Engineering & System Safety*, 153, 110–121, 2016.
- [21] Li, C., Mahadevan, S., (2016); Relative contributions of aleatory and epistemic uncertainty sources in time series prediction, *International Journal of Fatigue*, 82, 474–486, 2016.
- [22] Li, C.; Mahadevan, S. (2016); Role of calibration, validation, and relevance in multi-level uncertainty integration, *Reliability Engineering & System Safety*, 148, 32–43, 2016.
- [23] Liu, H. C., You, J. X., Fan, X. J., Lin, Q. L., (2014); Failure mode and effects analysis using d numbers and grey relational projection method, *Expert Systems with Applications*, 41 (10), 4670–4679, 2014.
- [24] Liu, H. C., You, J. X., Li, P., Su, Q., (2016); Failure mode and effect analysis under uncertainty: An integrated multiple criteria decision making approach, *IEEE Transactions on Reliability*, 65 (3), 1380–1392, 2016.
- [25] Liu, H. C., You, J. X., Li, P., Su, Q., (2016); Failure mode and effect analysis under uncertainty: An integrated multiple criteria decision making approach, *IEEE Transactions on Reliability*, 1–13, 2016.
- [26] Liu, H.-C., You, J.-X., You, X.-Y., Shan, M.-M., (2015); A novel approach for failure mode and effects analysis using combination weighting and fuzzy vikor method, *Applied Soft Computing*, 28, 579–588, 2015.
- [27] Liu, J., Lian, F., Mallick, M., (2016); Distributed compressed sensing based joint detection and tracking for multistatic radar system, *Information Sciences*, 369, 100–118, 2016.

- [28] Liu, T., Deng, Y., Chan, F., (2017); Evidential supplier selection based on DEMATEL and game theory, *International Journal of Fuzzy Systems*, DOI: 10.1007/s40815-017-0400-4, 2017.
- [29] Liu, W., Liu, H. B., Li, L. L., (2017); A multiple attribute group decision making method based on 2-d uncertain linguistic weighted heronian mean aggregation operator, *International Journal of Computers Communications & Control*, 12 (2), 254-264, 2017.
- [30] Liu, Z., Pan, Q., Dezert, J., Han, J.-W., He, Y., (2017); Classifier fusion with contextual reliability evaluation, *IEEE Transactions on Cybernetics*, 99, 1-14, 2017.
- [31] Liu, Z., Pan, Q., Dezert, J., Martin, A., (2017); Combination of classifiers with optimal weight based on evidential reasoning, *IEEE Transactions on Fuzzy Systems*, 99, 1-15, 2017.
- [32] Méndez-González, L. C., Ambrosio-Lazaro, R., Rodriguez-Borbon, I., Alvarado-Iniesta, A., (2016); Failure mode and effects analysis of power quality issues and their influence in the reliability of electronic products, *Electrical Engineering*, 1-13, 2016.
- [33] Mo, H., Deng, Y., (2016); A new aggregating operator in linguistic decision making based on d numbers, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 24 (6), 831-846, 2016.
- [34] Mo, H., Gao, C., Deng, Y., (2015); Evidential method to identify influential nodes in complex networks, *Journal of Systems Engineering and Electronics*, 26 (2), 381-387, 2015.
- [35] Mo, H., Lu, X., Deng, Y., (2016); A generalized evidence distance, *Journal of Systems Engineering and Electronics*, 27 (2), 470-476, 2016.
- [36] Perdomo Ojeda, M., Salomon Llanes, J., (2016); Expanded failure mode and effects analysis: Advanced approach for reliability assessments, *Revista Cubana de Ingenieria*, 7 (2), 5-14, 2016.
- [37] Pillay, A., Wang, J., (2003); Modified failure mode and effects analysis using approximate reasoning, *Reliability Engineering System Safety*, 79 (1), 69-85, 2003.
- [38] Schneider, H., (1996); Failure mode and effect analysis: Fmea from theory to execution, *Technometrics*, 38 (1), 80-80, 1996.
- [39] Tooranloo, H. S., Ayatollah, A. S., (2016); A model for failure mode and effects analysis based on intuitionistic fuzzy approach, *Applied Soft Computing*, 49, 238-247, 2016.
- [40] Viejo, M. R., Sanchez-Izquierdo Riera, J., Molano, E., Barea Mendoza, J. A., Temprano, V. S., Diaz, C. L., Montejo Gonzalez, J. C., (2016); Improvement of the safety of a clinical process using failure mode and effects analysis: Prevention of venous thromboembolic disease in critical patients, *Medicina Intensiva*, 40 (8), 483, 2016.
- [41] Wang, J., Hu, Y., Xiao, F., Deng, X., Deng, Y., (2016); A novel method to use fuzzy soft sets in decision making based on ambiguity measure and Dempster-Shafer theory of evidence: An application in medical diagnosis, *Artificial intelligence in medicine*, 69, 1-11, 2016.
- [42] Wang, J., Xiao, F., Deng, X., Fei, L., Deng, Y., (2016); Weighted Evidence Combination Based on Distance of Evidence and Entropy Function, *International Journal of Distributed Sensor Networks*, 12 (7), 2016.

- [43] Wang, Y. M., Chin, K. S., Poon, G. K. K., Yang, J. B., (2009); Risk evaluation in failure mode and effects analysis using fuzzy weighted geometric mean, *Expert Systems with Applications*, 36 (2), 1195–1207, 2009.
- [44] Xiao, F., Aritsugi, M., Wang, Q., Zhang, R., (2016); Efficient processing of multiple nested event pattern queries over multi-dimensional event streams based on a triaxial hierarchical model, *Artificial Intelligence in Medicine*, 72 (C), 56–71, 2016.
- [45] Xiao, F., Zhan, C., Lai, H., Tao, L., Qu, Z., (2017); New parallel processing strategies in complex event processing systems with data streams, *International Journal of Distributed Sensor Networks*, 13 (8), 1–15, 2017.
- [46] Xu, S., Jiang, W., Deng, X., Shou, Y., (2017); A modified physarum-inspired model for the user equilibrium traffic assignment problem, *Applied Mathematical Modelling*, In Press, DOI: 10.1016/j.apm.2017.07.032, 2017.
- [47] Yang, Z., Liu, P., Zhu, Y., Zhang, Y., (2016); A comprehensive reliability allocation method for series systems based on failure mode and effects analysis transformed functions, *Proc. of the Institution of Mechanical Engineers, Part B- Journal of Engineering Manufacture*, 230, 2239–2248, 2016.
- [48] Zhang, D., (2017); High-speed train control system big data analysis based on fuzzy rdf model and uncertain reasoning, *International Journal of Computers Communications & Control*, 12 (4), 577–591, 2017.
- [49] Zhang, Q., Li, M., Deng, Y., (2017); Measure the structure similarity of nodes in complex networks based on relative entropy, *Physica A: Statistical Mechanics and its Applications*, 10.1016/j.physa.2017.09.042, 2017.
- [50] Zhang, X., Deng, Y., Chan, F. T. S., Adamatzky, A., Mahadevan, S., (2016); Supplier selection based on evidence theory and analytic network process, *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 230 (3), 562–573, 2016.
- [51] Zhang, X., Mahadevan, S., (2017); A game theoretic approach to network reliability assessment, *IEEE Transactions on Reliability*, 66 (3), 875–892, 2017.
- [52] Zheng, H., Deng, Y., Hu, Y., (2017); Fuzzy evidential influence diagram and its evaluation algorithm, *Knowledge-Based Systems*, 131, 28–45, 2017.
- [53] Zheng, X., Deng, Y., (2017); Dependence assessment in human reliability analysis based on evidence credibility decay model and iowa operator, *Annals of Nuclear Energy*, accepted, 2017.
- [54] Zhou, X., Deng, X., Deng, Y., Mahadevan, S., (2017); Dependence assessment in human reliability analysis based on d numbers and ahp, *Nuclear Engineering and Design*, 313, 243–252, 2017.
- [55] Zhou, X., Shi, Y., Deng, X., Deng, Y., (2017); D-DEMATEL: A new method to identify critical success factors in emergency management, *Safety Science*, 91, 93–104, 2017.

# An Improved Local Descriptor based Object Recognition in Cluttered 3D Point Clouds

X. Liu, Y. Lu, T. Wu, T. Yuan

## Xiaoni Liu

Key Lab of Symbol Computation and  
Knowledge Engineer of Ministry of Education,  
College of Computer Science and Technology,  
Jilin University, Changchun 130012, China  
liu\_xiaoni@foxmail.com

## Yinan Lu\*

Key Lab of Symbol Computation and  
Knowledge Engineer of Ministry of Education,  
College of Computer Science and Technology,  
Jilin University, Changchun 130012, China  
\*Corresponding author: luyn@jlu.edu.cn

## Tieru Wu

College of Mathematics,  
Jilin University, Changchun 130012, China  
wutr@jlu.edu.cn

## Tianwen Yuan

College of Computer Science and Technology,  
Jilin University, Changchun 130012, China  
yuantw14@mails.jlu.edu.cn

**Abstract:** Object recognition in three-dimensional point clouds is a new research topic in the field of computer vision. Numerous nuisances, such as noise, a varying density, and occlusion greatly increase the difficulty of 3D object recognition. An improved local feature descriptor is proposed to address these problems in this paper. At each feature point, a local reference frame is established by calculating a scatter matrix based on the geometric center and the weighted point-cloud density of its neighborhood, and an improved normal vector estimation method is used to generate a new signature of histograms of orientations (SHOT) local-feature descriptor. The geometric consistency and iterative closest point method realize 3D model recognition in the point-cloud scenes. The experimental results show that the proposed SHOT feature-extraction algorithm has high robustness and descriptiveness in the object recognition of 3D local descriptors in cluttered point-cloud scenes.

**Keywords:** 3D point cloud, local feature, object recognition, noise, density variation.

## 1 Introduction

With the development of computer software theory and hardware, three-dimensional (3D) model data have become the fourth-generation multimedia data type after sound, images, and video. The 3D point cloud is an important representation of 3D model data. The purpose of object recognition in 3D point clouds is to identify the object of interest from the 3D point-cloud scenes and obtain its pose (such as position and orientation). Object recognition in 3D point clouds has become a research hotspot in the fields of interior decoration, robot arm crawling, urban planning, marine environmental monitoring, computer-aided design and manufacturing, biomedicine, smartphones, and aerospace technology.

Feature extraction is a very critical step and affects the performance of identification systems. The feature descriptors can be categorized into two groups: global and local descriptors [10]. Global descriptors compute the feature for a cluster of points representing a whole object but rather for individual points. Hence, they encode the geometry of the whole object using their  $k$ -neighborhood data. This has led to promising results in 3D model retrieval and shape classification. A pre-segmentation step is required before the computation of the global descriptors of a given object, in order to obtain the cluster that represents the object. The most popular global descriptors are the viewpoint feature histogram (VFH) [1], clustered VFH (CVFH) [2], ensemble of shape functions (ESF) [24], and global fast PFH [21]. Local descriptors are computed for individual points by using its underlying geometrical information, obtained from the query point  $k$ -neighborhood. They include point signatures (PSs) [7], spin images (SIs) [14], 3D shape context (3DSC) [8], the point feature histogram (PFH) [19], the fast PFH (FPFH) [20], the signature of histograms of orientations (SHOT) [23], the radius-based surface descriptor [3], the unique-shape context descriptor [11], the rotation-invariant feature transform [15], and the normal-aligned radial feature [6].

For a qualified local surface feature, robustness and descriptiveness are very important factors. A histogram is used to accurately describe the local area, when especially considering the spatial distribution of the neighboring points.

It is a challenging task to achieve high descriptiveness and strong robustness to various nuisances faced by the existing methods. SHOT can compete against most descriptors and can gather more information. It has two disadvantages: a large amount of computational storage is required, and the original LRF of SHOT is sensitive to the point-cloud density variation. An improved SHOT descriptor is proposed to address the noise and varying point cloud density problems in this paper. An improved scatter matrix computation based on weighted distance and point-cloud density is used for establishing an LRF and generating the normal for each point in SHOT descriptor. In an LRF, the position of geometric center is iteratively calculated based on the weighted point-cloud density. The experimental results show that the proposed feature-extraction method is robust to noise and point-cloud variations, and highly efficient for the 3D object recognition in cluttered point-cloud scenes.

The remainder of this paper is organized as follows. Section 2 briefly reviews previous studies related to point-cloud descriptors. Section 3 describes the proposed SHOT feature and a 3D object-recognition algorithm. Sections 4 and 5 present the feature-matching results and the object-recognition results for a standard dataset, respectively. Finally, we draw conclusions in Section 6.

## 2 Related work

Numerous studies on mesh and point clouds have been performed, and the methods based on mesh algorithms have achieved accurate results. SI is a famous local surface feature. Johnson et al. got the normal of a feature point to form the corresponding axis and computed 2D histograms according to two parameters, finding that the SI is robust to occlusion and clutter. To achieve a high descriptiveness, Frome et al. [8] proposed a 3DSC descriptor, as an extension of the 2D shape context method. It considered a sphere superimposed on the query point and divided the radius neighbors into smaller bins, counted the points in each bin and weighted inversely with the density of the bin. Guo et al. [9] proposed a method for constructing the LRF and extracting the rotational projection statistics (RoPS) descriptor by projecting and quantizing the rotated neighbors on the 2D  $xy$ ,  $xz$ , and  $yz$  planes. The readers may refer to a recent survey [10] for additional details.

Owing to the unprecedented levels of computing science, Aldoma and Marton et al. [1] introduced the VFH. Aldoma and Tombari et al. [2] proposed a method called as CVFH. Wohlkinger et al. [24] collected ten shape functions, which depicting the point cloud, which is called the ESF. The VFH, CVFH, and ESF descriptors all use normal method. Compared with the moderate processing time of the VFH and ESF, that of the CVFH is very high.

When the defined search radius only covers the neighborhood of the keypoint, the PFH can be used as a local descriptor in [22]. The geometrical properties of each target point were estimated by analyzing the difference between the directions of the normal in the  $k$ -neighborhood of the keypoint and building a multidimensional histogram of values. One drawback of the PFH is its high computational complexity. Rusu et al. [20] also proposed an FPFH descriptor, which was a simplified and modified version of the PFH that used less neighbors for counting the orientation diversities. Compared with the PFH, the FPFH was designed to reduce the computational complexity from  $O(nk^2)$  to  $O(nk)$  while maintaining similar descriptiveness. Further, Zhong [27] developed the intrinsic shape signature (ISS) based on the radius, elevation, and azimuth angle; the support domain was divided into several units of the same size, and the multidimensional ISS feature description was formed by counting the number of neighbors in each unit. The ISS outperformed the SI and 3DSC in the model-retrieval aspects. Tombari et al. [23] presented a SHOT local 3D descriptor that is effective and efficient for surface-matching applications. It is designed to lie at the intersection between the two classes gathering the vast majority of data in the area of 3D description, i.e., signatures and histograms. SHOT outperformed state-of-the-art local descriptors in many experiments addressing descriptor matching for object recognition, 3D reconstruction, and shape retrieval.

For local feature-based 3D object recognition in the point cloud, there involve three main steps: feature matching, hypothesis generation, and hypothesis verification. Chua and Jarvis [7] employed a PS descriptor of a scene to match against each model and then used pairs of corresponding points to calculate the rigid transformation between a candidate model and a specified scene, at the cost of considerable memory and time. Johnson and Hebert [14] matched the SI descriptor of a scene with their models to generate point correspondences, which they grouped using the hypothesis generation of geometric consistency (GC). This algorithm is not only robust to clutter and occlusion but also has robustness to complicated real scenes. Yamany and Farag [25] proposed a feature descriptor for surface signatures. Mian et al. [17] used a tensor descriptor to obtain the feature correspondences and a hypothesis model. The hypothesis model was then transformed into the scene, and the hypothesis was verified using the iterative closest point (ICP) algorithm [4]. This method is superior to the SI-based algorithm with regard to the recognition rate and efficiency. Mian et al. [18] developed a 3D object-recognition algorithm based on keypoint matching, which was used to recognize the different and unknown scales of objects.

### 3 SHOT-based 3D object recognition

#### 3.1 SHOT feature

A 3D local feature should be invariant for rigid transformations that include rotation and translation. Thus, we establish a single and repeatable improved LRF to represent the local surface. A local feature of a 3D point cloud should have a very high descriptiveness to ensure the accuracy and timeliness of the object recognition. An improved normal vector estimation method is used to generate a new SHOT local-feature description.

Guo et al. [12] proposed that the singleness and repeatability of the LRF directly affect the follow-up descriptor descriptiveness and robustness. We can see that the construction of the LRF

is essential. The original SHOT feature is robust and descriptive to noise, object occlusion, and complex scenes, but the robustness to point-cloud variations is poor. Therefore, an improved LRF is proposed to solve the problem that the original SHOT feature is sensitive to changes in the point-cloud density.

The original SHOT typically uses three orthogonal eigenvectors  $V = \{v_1, v_2, v_3\}$  obtained via eigenvalue decomposition (EVD) of the scatter matrix to represent three coordinate axes for establishing the LRF and only considers the noise in order to add the distance weight. It then directly calculates the feature point  $p$  to replace the center point  $\tilde{p}$ . The scatter matrix of the original SHOT is given as follows:

$$M = \frac{1}{\sum_{i:d_i \leq R} (R - d_i)} \sum_{i=0}^k (R - d_i)(p_i - p)(p_i - p)^T \quad (1)$$

Where  $R$  represents the radius of the support fields, and  $d_i = \|p_i - p\|$  represents the Euclidean distance between the supporting neighborhood and the feature point. The original weighted point-cloud density and the improved weighted point-cloud density are showed below, where  $p_i$  is the feature point;  $p_j$  is the supporting neighborhood point;  $\tilde{p}$  is the geometric center in different cloud densities, the selection of  $\beta$  will be described in detail in Section 4.3:

$$W_{ij} = \frac{1}{(R - \|p_j - \tilde{p}_i\|)} \quad (2)$$

$$W'_{ij} = \frac{W_{ij}}{\text{density}^\beta(p_j)} = \frac{1}{(R - \|p_j - \tilde{p}_i\|) * \text{density}^\beta(p_j)} \quad (3)$$

The scatter matrix for establishing the LRF is calculated as follows:

$$COV(p_i) = \frac{\sum_{\|p_j - p_i\| \leq R} W'_{ij} \cdot (p_j - p_i)(p_j - p_i)^T}{\sum_{\|p_j - p_i\| \leq R} W'_{ij}} \quad (4)$$

On the other hand, to maintain consistency with the original SHOT feature, we improve Lin's method [16] by means of establishing the scatter matrix to calculate the center point at different cloud densities. The improved center point is designated as the geometric center, requiring that the distance between all points in the supporting area and the geometric center, which calculated as below formula (5) is the smallest, where the distance refers to the weighted Euclidean distance. Originally, it is very sensitive to calculate the center of the scatter matrix when the point-cloud density changes, however, the effect of the geometric center on this matter is very small.

$$\arg \min \sum_{\|p_j - p_i\| < R} W'_{ij} \cdot \|p_j - \tilde{p}_i\| \quad (5)$$

The algorithm is described as follows.

Algorithm: Compute an geometric center in different cloud densities  $\tilde{p}_i$

Input: the neighborhoods  $P = \{p_1, p_2, \dots, p_n\}$  of the feature point  $p_i$  within the support radius and the threshold  $\varepsilon$

Output: geometric center  $\tilde{p}_i$

1. Initialization:  $\tilde{p}_i = p_i$

2.  $p_0 = \tilde{p}_i$

$$W'_{ij} = \frac{1}{(R - \|p_j - \tilde{p}_i\|) * \text{density}^\beta(p_j)}$$

3. Update the geometric center and the weight

$$\tilde{p}_i = \frac{\sum_{j=1}^n W_{ij} * p_j}{\sum_{j=1}^n W_{ij}}$$

4. If  $\|\tilde{p}_i - p_0\| < \varepsilon$ , output  $\tilde{p}_i$ ; otherwise, go to step 2

When the EVD is performed by the scatter matrix, the eigenvector direction is both positive and negative. Sign ambiguity is generated when an LRF is established directly by using the characteristic direction of the EVD. To solve this problem, SHOT employs disambiguated EVD [5] to make all the neighbors in the support to the center point as normal as possible.

Take the  $x$  axis as an example.

$$\begin{aligned} S_x^+ &= \{i | d_i \leq R \&\& (p_i - p) \cdot x^+ \geq 0\} \\ S_x^- &= \{i | d_i \leq R \&\& (p_i - p) \cdot x^- > 0\} \end{aligned} \quad (6)$$

Considering the same number of  $S_x^+$  and  $S_x^-$ , we first select the nearest  $k$  points around the feature point  $P_i$  ( $k = 11$ ) and then count their  $\tilde{S}_x^+$  and  $\tilde{S}_x^-$ :

$$\begin{aligned} \tilde{S}_x^+ &= \{i | i \in M(k) \&\& (p_i - p) \cdot x^+ \geq 0\} \\ \tilde{S}_x^- &= \{i | i \in M(k) \&\& (p_i - p) \cdot x^- > 0\} \end{aligned} \quad (7)$$

The final positive and negative directions of the  $x$  axis are determined by the number of statistics:  $S_x^+$ ,  $S_x^-$ ,  $\tilde{S}_x^+$ ,  $\tilde{S}_x^-$ .

$$x = \begin{cases} x^+, |S_x^+| > |S_x^-| \\ x^-, |S_x^+| < |S_x^-| \\ x^+, |S_x^+| = |S_x^-| \&\& |\tilde{S}_x^+| > |\tilde{S}_x^-| \\ x^-, |S_x^+| = |S_x^-| \&\& |\tilde{S}_x^+| < |\tilde{S}_x^-| \end{cases} \quad (8)$$

The signature of SHOT is calculated as the cosine of the angle between the normal at the point  $n_q$  and the local  $Z$  axis at the feature point  $p$  (introducing bin by the radius, the horizontal angle and the elevation division, counting the established histogram by the angle between normal vector of neighborhood and  $Z$  axis). However, in this paper, the normal of the neighborhood has not been used, but in accordance with the above calculation of the LRF method to recalculate the scatter matrix of distance weight and point-cloud density weight for each point, with the minimum eigenvalue corresponding to the feature direction to replace the normal:

$$\cos \theta = n'_q \cdot Z_p \quad (9)$$

where  $n'_q$  is the minimum eigenvalue corresponding to the feature direction, which is calculated using the improved scatter matrix and the geometric center, and  $Z_p$  is the  $Z$ -axis direction of the LRF.

### 3.2 3D object recognition

The proposed 3D object-recognition algorithm follows the framework of the object-recognition scheme based on local features, including feature-point detection, feature-description extraction, feature matching, hypothesis generation, and hypothesis verification. A block diagram of the proposed algorithm is shown in Figure 1. Then we introduce four main modules.

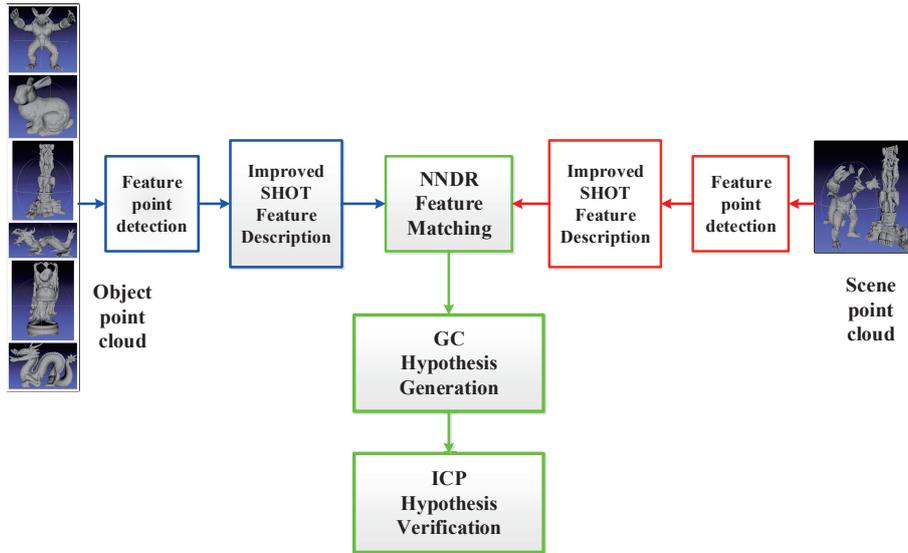


Figure 1: Block diagram of the object-recognition algorithm for 3D point-cloud scenes.

Module 1. Feature-point detection: We use the ISS method with boundary-point removal (ISS-BR) for feature-point detection in Point Cloud Library (PCL) [28] to extract the feature points in the object and scenes.

Module 2. Feature-description extraction: For each feature point, we define an LRF to generate a SHOT feature.

Module 3. Feature matching: We employ the statistical strategy of the nearest-neighbor distance ratio (NNDR); that is, when feature ratio of an object is lower than the given threshold, the two features are considered a match. The position of the object in the scenes is estimated according to the GC at the hypothesis-generation stage.

Module 4. Hypothesis verification: Because the hypothesis-generation phase can only be produced a preliminary coarse transformation, it is necessary to perform hypothesis verification which separates the correct or error hypothesis, so as to accurately calculate the rotation and translation transformation matrix from object to scene. We use the ICP hypothesis-validation method [26].

## 4 Evaluation of SHOT feature

The experimental results show that LFR in this paper has strong robustness to different point-cloud density, and our SHOT feature significantly outperforms the original SHOT features with regard to recall and precision and it also has a higher robustness to noise and point-density variation. Compared with the FPFH feature, our SHOT feature exhibits the best performance under all the levels of mesh decimation and noise.

### 4.1 Robustness to LRF

Regarding robustness, Guo et al. [9] proposed a criteria for the evaluation of the LRF. For a pair of corresponding feature points  $(k_M, k_S)$ , the established LRF is a  $3 \times 3$  matrix  $(L_M, L_S)$  that represents the directions of the three axes  $XYZ$ , and the error between the two LRFs can be expressed as

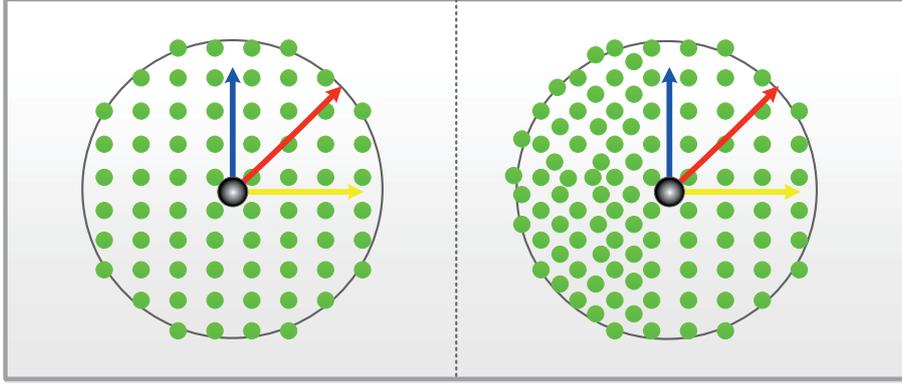


Figure 2: Established LRF for different point clouds.

$$\varepsilon = \arccos \left( \frac{\text{trace}(L_M L_S^{-1}) - 1}{2} \right) \frac{180}{\pi} \quad (10)$$

where *trace* indicates the trajectory of the matrix, which is the sum of the elements on the main diagonal. If the two LRFs are the same,  $L_M L_S^{-1}$  is a unit matrix,  $\text{trace}(L_M L_S^{-1})$  is 3, and the final result  $\varepsilon$  is 0. Similarly,  $\varepsilon$  can be denoted as the angle of rotation error between the two LRFs; a smaller  $\varepsilon$  yields a smaller error.  $\varepsilon$  indicates that there is no error.

## 4.2 Performance for the retrieval dataset

We use the retrieval dataset [29] to test the performance of our SHOT feature. It is well known that the retrieval dataset is a famous real-life datasets. It contains six models (Armadillo, Bunny, Thai Statue, Happy Buddha, Dragon, and Asian Dragon) and 18 scenes which taken from the Stanford 3D Scanning Repository [30]. And for creating clutter and pose variances, six models are randomly putted in a scene.

Each scene is divided into three levels of noise with standard deviations of 0.1mr, 0.3mr, and 0.5mr (where mr refers to the mesh resolution, i.e., the average distance of each point to the nearest point). Both the model and the scene are stored in the Polygon File Format (PLY) mesh model, and the transformation matrix for each model to the 3D scene is stored in the accompanying XF file. Armadillo model of the retrieval dataset and three noise scenes are shown in Figure 3. Four sample models are shown in Figure 4.

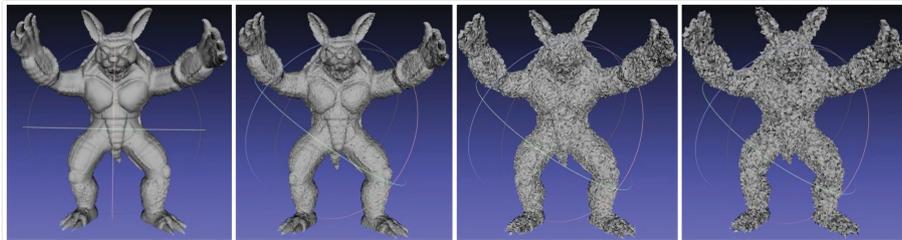


Figure 3: Armadillo model of the retrieval dataset and three noise scenes.

We get 1,000 feature points from each model to evaluate the robustness of the LRF, meanwhile, the corresponding of 1,000 feature points are also sampled to generate ground-truth correspondences. And we resample each scene to 1/2, 1/4, and 1/8 of its original mesh resolution,



Figure 4: Four sample models of the Stanford 3D Scanning Repository.

adding the three aforementioned noise scenes. Simultaneously, we select three types of LRFs - a single LRF, the LRF of the SHOT descriptor, and the LRF of the improved SHOT, which are tested with varying mesh resolutions and noise levels.

From Figure 5 we can see the average percentage of the LRF angle errors for five scenes and models at the three point-cloud densities and noise intensities. The abscissa shows the angular-error interval; e.g., 10 represents  $[0, 10)$ . For convenient statistics, we don't consider the case that the LRF angle error is greater than 90 degrees. Figure 5 shows that, for a noise with a standard deviation of 0.1mr and 0.3mr, the improved SHOT one is more stable than the original SHOT one, in particular, the original SHOT LRF has the best performance in the error interval  $[0, 10]$ , but drops to the end in the error interval  $[10, \infty)$ . On the other hand, the improved SHOT one is more strong to noise than the original SHOT one for a noise with a standard deviation of 0.5mr.

### 4.3 Descriptiveness assessment

We use the precision-recall curve (PRC), which comprises pairs of values of the recall and precision, to evaluate the descriptiveness of the 3D local-feature descriptor. The PRC is generated as follows. First, numerous keypoints are extracted from the 3D point cloud corresponding to the scene and all the models; the feature descriptors of these keypoints are then computed. Second, to determine whether the match is robust, the NNDR technique is used to perform feature matching between the two descriptors. For each feature, the match is regarded as robust if the condition  $\|N_s - N_M\|/\|N_s - N'_M\| \leq \tau$  is met, where  $N_s$  is the descriptor in the scene, and  $N_M$  and  $N'_M$  are its nearest and second-nearest neighbor in the model. The *precision* is calculated as the number of correct matches with respect to the total number of matches:

$$precision = \frac{\text{The number of correct matches}}{\text{The number of matches}} \quad (11)$$

The *Recall* is calculated as the number of correct matches with respect to the number of corresponding features between the scene and models:

$$Recall = \frac{\text{The number of correct matches}}{\text{The number of corresponding features}} \quad (12)$$

Where the value of the threshold  $\tau$  is changed from 0 to 1. To test the descriptiveness of the descriptors, the ISS-BR was used to detect the keypoints in the scene and the models.

For testing the capability of the original SHOT, our improved SHOT, and the FPFH with varying mesh resolutions, we resample each scene to 1/2, 1/4, and 1/8 of its original mesh resolution. On the other hand, for evaluating the robustness to noise, we choose noise with standard deviations of 0.1mr, 0.3mr, and 0.5mr to add to each scene. The feature-point support field is set to six times the average mesh resolution, and the RPC results at different levels of the mesh resolution and noise are shown in Figure 6.

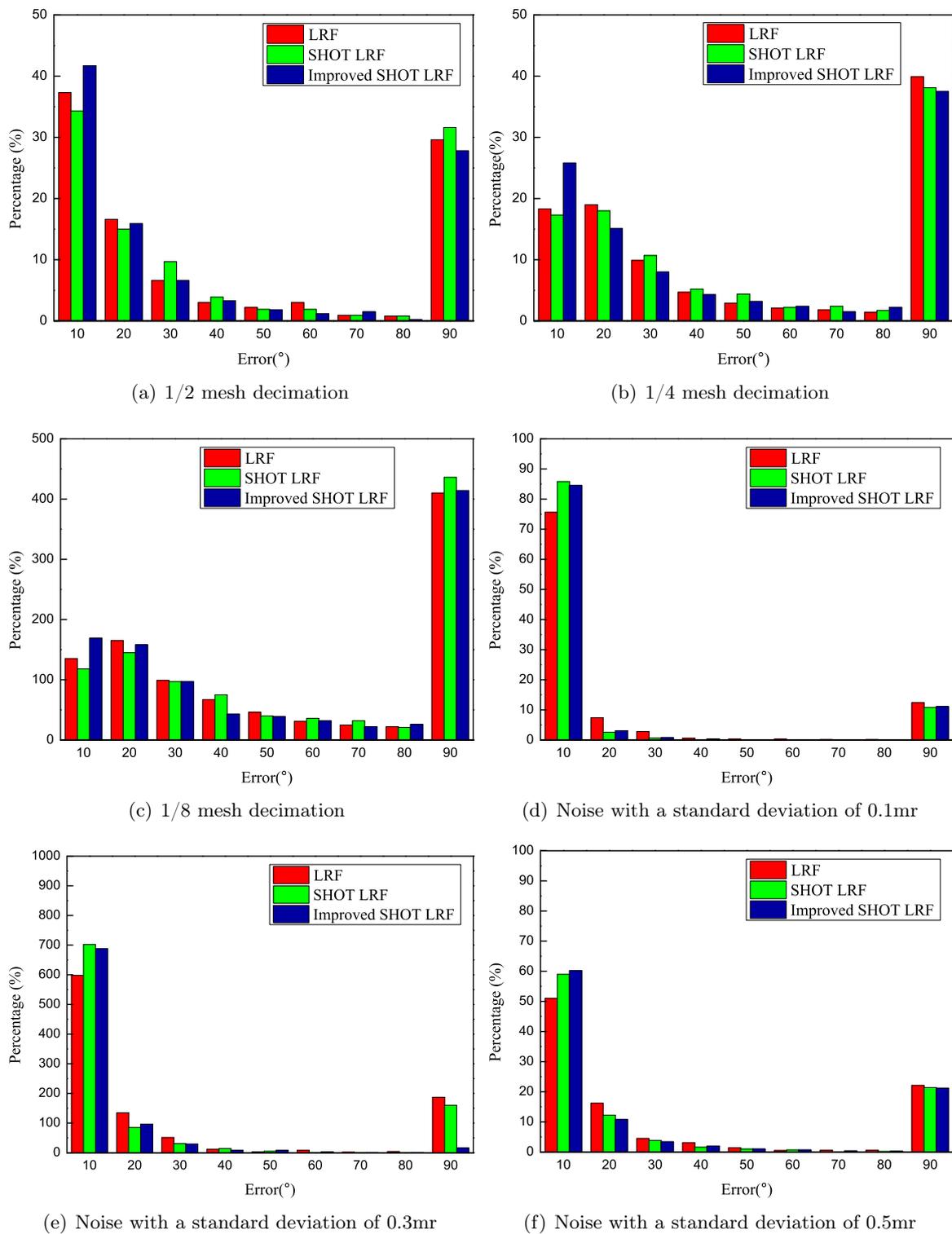


Figure 5: Comparison of different LRFs

As you can see from the graph, for varying mesh resolutions and noise levels, the improved SHOT was more descriptive and robust. Regarding the robustness to varying mesh resolutions, our SHOT feature significantly outperform the other features, and the original SHOT is the

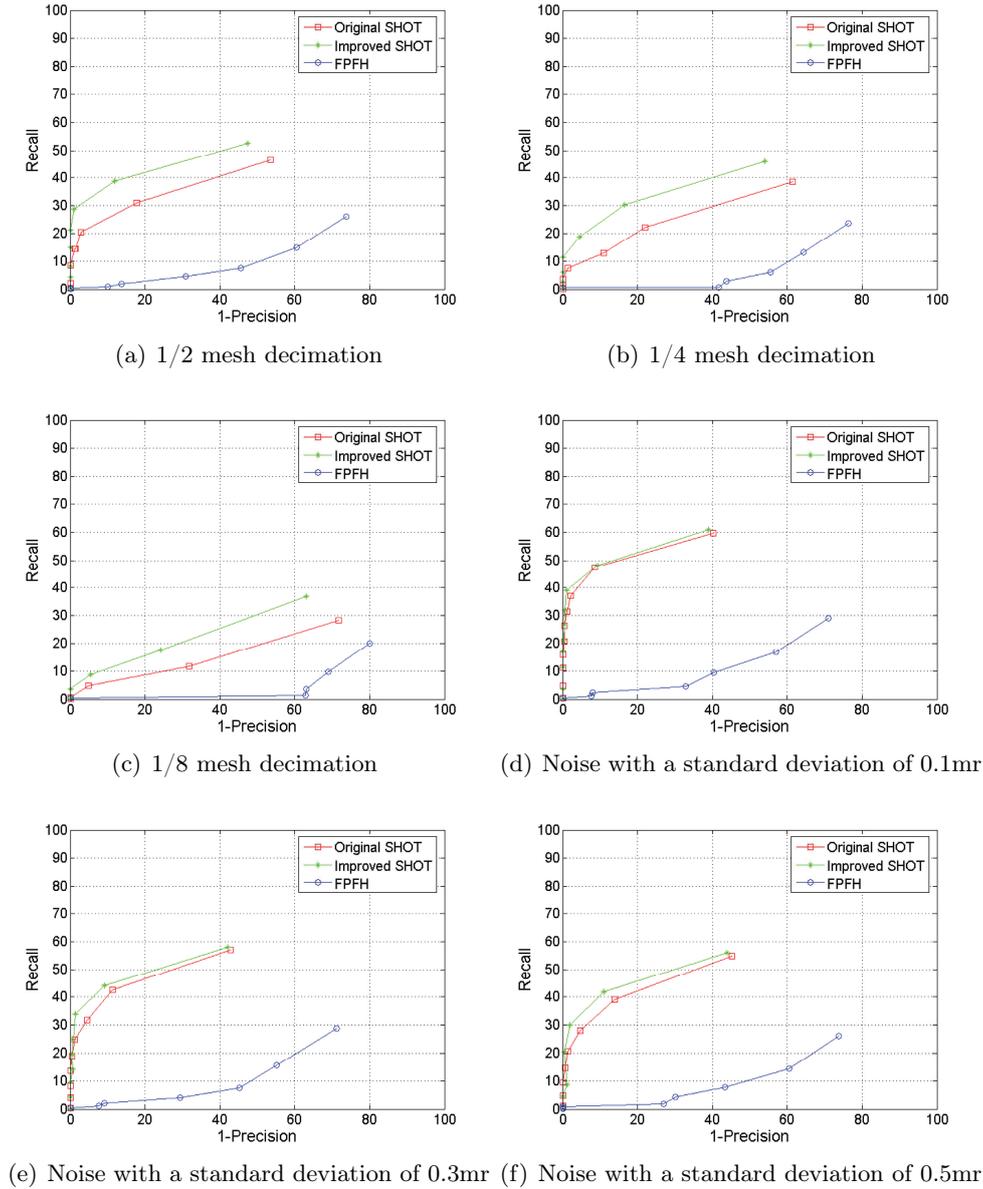


Figure 6: Recall vs. 1-precision curves with respect to the mesh resolution and noise

second. The performance of the improved SHOT feature under 1/4 and 1/8 mesh decimations is better than that of the FPFH, which is unsatisfactory. Regarding the robustness to noise, our SHOT feature exhibits the best force at all noise levels.

In Figure 7, we also show the performance of transforming the point-cloud density into a  $\beta$  square. In the case of resampling the original mesh resolution of 1/2 and noise with standard deviations of 0.1mr, the best results are obtained with  $\beta = 1$ .

## 5 Evaluation of the 3D object-recognition algorithm

To examine the performance of the 3D object-recognition algorithm with our improved SHOT feature, we perform an object-recognition experiment using the BoD1 dataset [31]. We use the improved SHOT in our object-recognition algorithm to perform experiments on this dataset using

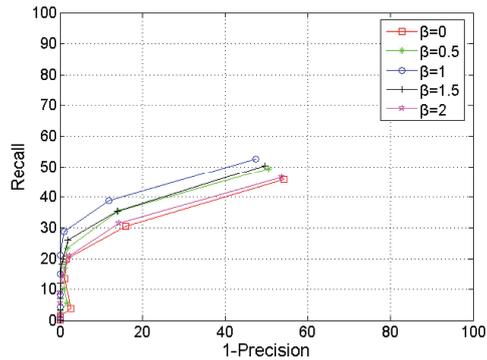


Figure 7: Performance of the SHOT features with different values of  $\beta$ .

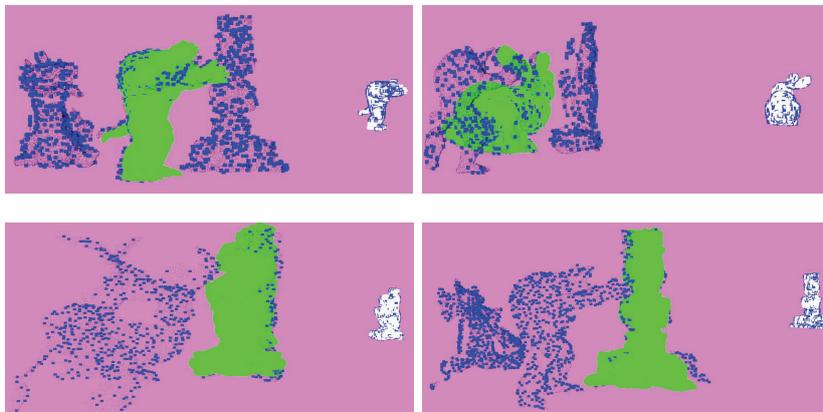


Figure 8: Correct recognition results of our algorithm

an Intel(R) Core(TM) i7-6500K 3.20-GHz Windows machine with 16 GB of RAM.

In this study, all the models from the BoD1 dataset with noise having a standard deviation of 0.1mr are resampled with the original mesh resolution of 1/2 as the experimental model library. In the recognition phase, we perform a comparative experiment with the original SHOT, our improved SHOT, and the FPFH, where we use 10 times the average mesh resolution as the radius of the calculation normal and 15 times the average mesh resolution as the radius of SHOT and FPFH descriptors. According to the descriptive experiment, an NNDR value of 0.9 was more appropriate.

Figure 8 shows the correct results of recognizing the object in the 3D cluttered point-cloud scenes via ISS-BR feature point detection with our improved SHOT feature. The black points represent the scene models of the original mesh resolution of 1/2, with added noise having a standard deviation of 0.1mr. The original object is represented by the white models on the left, and the blue points represent the feature points detected from the model and scenes. The green model indicates the locations of the objects in the scenes.

The recognition rate [13] is the ratio of the number of correct recognition to the total number of experiments:

$$\text{recognition rate} = \frac{\text{The number of correct recognition}}{\text{The total number of experiments}} \quad (13)$$

For a rigorous comparison, we perform 3D object-recognition experiments on a subset of BoD1, which contained 24 scenes of Armadillo and Bunny, 18 scenes of Dragon and Happy Buddha, and 12 scenes of Asian Dragon and Thai Statue. The recognition rates are presented

Table 1: Recognition rates for a subset of the BoD1 dataset.

Algorithm(%)	Armadillo	Bunny	Dragon	Asian Dragon	Happy,Buddha	Thai,Statue
Improved SHOT	66.7	83.3	66.7	83.3	55.6	100
SHOT	54.2	83.3	55.6	83.3	50	100
FPFH	41.7	75	44.4	75	38.9	50

in Table 1. The improved SHOT algorithm exhibits a higher recognition rate by 20.4% than the FPFH-based algorithm. The improved SHOT algorithm also outperforms the original SHOT algorithm, by 74.6% and 71.1%. The improved SHOT algorithm exhibits the highest recognition rate for all the models and achieves a recognition rate of 100% for the Thai Statue model.

The experimental results are analyzed as follows.

The original LRF of SHOT is sensitive to the point-cloud density; thus, simultaneously improving the normal and LRF to achieve the performance is more obvious. The performance of SHOT is better than that of FPFH; thus, the SHOT recognition performance is better than that of the FPFH.

Given these considerations, SHOT descriptors are the best option for time-crucial applications and for applications requiring high levels of descriptiveness and efficiency. On the other hand, for space-crucial applications, FPFH is better because of its low memory requirements for the feature storage.

Despite its benefits, the SHOT feature has disadvantages. It is designed for fixed-scale objects; once the object changes scale, the SHOT feature may face challenges. Therefore, to solve the problem of non-rigid change, future studies should improve the feature description so that it has a high degree of robustness for not only noise, varying density, and occlusion but also non-rigid changes.

## 6 Conclusion

We propose a new SHOT feature for local description. At each feature point, an LRF is established by calculating a scatter matrix according to the geometric center and the weighted point-cloud density of its neighborhood. An improved normal vector estimation method is used to generate a new SHOT local-feature descriptor. The GC and the ICP method are used to perform 3D model recognition for the point-cloud scenes. The experimental results show that the proposed SHOT feature-extraction algorithm is robust to noise and point-cloud variations. Moreover, it is highly efficient and effective for the 3D object recognition in cluttered point-cloud scenes. In recent years, Deep learning automatically learns a feature representation of 3D data. We will combine effective handcrafted 3D features with deeply learned features to further improve the performance in the future.

## Acknowledgments

This research is supported by the Specialized Research Funds for the Doctoral Program of Higher Education, China (No.20130061110054) and the National Natural Science, China (NSF61373003).

## Bibliography

- [1] Aldoma, A.; Marton, Z.; Tombari, F.; Wohlkinger, W. (2012); Tutorial: Point Cloud Library: Three-Dimensional Object Recognition and 6 DOF Pose Estimation, *Robotics and Automation Magazine IEEE*, 19, 80-91, 2012.
- [2] Aldoma, A.; Tombari, F.; Rusu, R.; Vincze, M. (2012); OUR-CVFH - Oriented, Unique and Repeatable Clustered Viewpoint Feature Histogram for Object Recognition and 6DOF Pose Estimation, *LNCS*, 7476, 113-122, 2012.
- [3] Arbeiter, G.; Fuchs, S.; Bormann, R.; Fischer, J.; Verl, A. (2012); Evaluation of 3D feature descriptors for classification of surface geometries in point clouds, *Proceedings of IROS*, 1644-1650, 2012.
- [4] Besl, P.; McKay, N. (1992); A method for registration of 3-D shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14, 239-256, 1992.
- [5] Bro, R.; Acar, E.; Kolda, T. (2008); Resolving the sign ambiguity in the singular value decomposition, *Journal of Chemometrics*, 22, 135-140, 2008.
- [6] Ceron, A.; Prieto, F. (2013); Evaluating and Comparing of 3D Shape Descriptors for Object Recognition, *LNCS*, 8034, 484-492, 2013.
- [7] Chua, C.; Jarvis, R. (1997); Point signatures: a new representation for 3D object recognition, *International Journal of Computer Vision*, 25, 63-85, 1997.
- [8] Frome, A.; Huber, D.; Kolluri, R.; Blow, T.; Malik, J. (2004); Recognizing Objects in Range Data Using Regional Point Descriptors, *Lecture Notes in Computer Science*, 3023, 224-237, 2004.
- [9] Guo, Y.; Sohel, F.; Bennamoun, M.; Lu, M.; Wan, J. (2013); Rotational projection statistics for 3d local surface description and object recognition, *International Journal of Computer Vision*, 105, 63-86, 2013.
- [10] Guo, Y.; Bennamoun, M.; Sohel, F.; Lu, M.; Wan, J. (2014); 3D Object Recognition in Cluttered Scenes with Local Surface Features: A Survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36, 2270-2287, 2014.
- [11] Guo, Y.; Bennamoun, M.; Sohel, F.; Lu, M.; Wan, J. (2014); Performance Evaluation of 3D Local Feature Descriptors, *Proceedings of ACCV*, 178-194, 2014.
- [12] Guo, Y.; Sohel, F.; Bennamoun, M.; Lu, M.; Wan, J. (2013); TriSI: A Distinctive Local Surface Descriptor for 3D Modeling and Object Recognition, *Proceedings of GRAPP*, 86-93, 2013.
- [13] Guo, Y.; Sohel, F.; Bennamoun, M.; Wan, J.; Lu, M.; A Novel Local Surface Feature for 3D Object Recognition under Clutter and Occlusion, *Information Sciences*, 293, 196-213, 2014.
- [14] Johnson, A.; Hebert, M. (1999); Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21, 433-449, 1999.
- [15] Liao, S.; Chung, A.C.S. (2010); A novel multi-layer framework for non-rigid image registration, *Proceedings of ISBI*, 344-347, 2010.

- 
- [16] Lin, C.; Chen, J.; Su, P.; Chen, C. (2014); Eigen-feature analysis of weighted covariance matrices for LiDAR point cloud classification, *ISPRS Journal of Photogrammetry and Remote Sensing*, 94, 70-79, 2014.
- [17] Mian, A.; Bennamoun, M.; Owens, R. (2006); Three-dimensional model-based object recognition and segmentation in cluttered scenes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28, 1584-1601, 2006.
- [18] Mian, A.; Bennamoun, M.; Owens, R. (2010); On the Repeatability and Quality of Keypoints for Local Feature-based 3D Object Retrieval from Cluttered Scenes, *International Journal of Computer Vision*, 89, 348-361, 2010.
- [19] Rusu, R.; Blodow, N.; Marton, Z.; Beetz, M. (2008); Aligning point cloud views using persistent feature histograms, *Proceedings of IROS*, 3384-3391, 2008.
- [20] Rusu, R.; Blodow, N.; Beetz, M. (2009); Fast Point Feature Histograms (FPFH) for 3D registration, *Proceedings of ICRA*, 3212-3217, 2009.
- [21] Rusu, R.; Holzbach, A.; Beetz, M.; Bradski, G. (2009); Detecting and segmenting objects for mobile manipulation, *Proceedings of ICCVW*, 47-54, 2009.
- [22] Rusu, R. (2010); Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments, *KI - Künstliche Intelligenz*, 24, 345-348, 2010.
- [23] Tombari, F.; Salti, S.; Stefano, L. (2010); Unique signatures of histograms for local surface description, *Proceedings of ECCV*, 356-369, 2010.
- [24] Wohlkinger, W.; Vincze, M. (2011); Ensemble of shape functions for 3D object classification, *Proceedings of IEEE RO-BIO*, 2987-2992, 2011.
- [25] Yamany, SM.; Farag, AA. (2002); Surface signatures: an orientation independent free-form surface representation scheme for the purpose of objects registration and matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 1105-1120, 2002.
- [26] Zhang, Z. (1994); Iterative point matching for registration of free-form curves and surfaces, *International Journal of Computer Vision*, 13, 119-152, 1994.
- [27] Zhong, Y. (2009); Intrinsic shape signatures: A shape descriptor for 3D object recognition, *Proceedings of ICCVW*, 689-696, 2009.
- [28] PCL; [pointclouds.org/](http://pointclouds.org/).
- [29] 3D Keypoint Detection Benchmark; [www.vision.deis.unibo.it/keypoints3d/](http://www.vision.deis.unibo.it/keypoints3d/).
- [30] The Stanford 3D Scanning Repository; <http://www.vision.deis.unibo.it/keypoints3d/>.
- [31] SHOT: Unique Signatures of Histograms for Local Surface Description; [www.vision.deis.unibo.it/research/80-shot](http://www.vision.deis.unibo.it/research/80-shot).

# Estimating Warehouse Rental Price using Machine Learning Techniques

Y. Ma, Z. Zhang, A. Ihler, B. Pan

## Yixuan Ma

1. International Center for Informatics Research  
Beijing Jiao Tong University, China  
2. Department of Computer Science  
University of California Irvine, USA  
mayixuan@bjtu.edu.cn

## Zhenji Zhang\*

International Center for Informatics Research  
Beijing Jiao Tong University  
100044 No.3 Shangyuancun, Haidian, Beijing, China  
\*Corresponding author: zhjzhang@bjtu.edu.cn

## Alexander Ihler

Department of Computer Science  
University of California Irvine, USA  
ihler@ics.uci.edu

## Baoxiang Pan

Department of Civil and Environmental Engineering  
University of California Irvine, USA  
baoxiangp@uci.edu

**Abstract:** Boosted by the growing logistics industry and digital transformation, the sharing warehouse market is undergoing a rapid development. Both supply and demand sides in the warehouse rental business are faced with market perturbations brought by unprecedented peer competitions and information transparency. A key question faced by the participants is how to price warehouses in the open market. To understand the pricing mechanism, we built a real world warehouse dataset using data collected from the classified advertisements websites. Based on the dataset, we applied machine learning techniques to relate warehouse price with its relevant features, such as warehouse size, location and nearby real estate price. Four candidate models are used here: Linear Regression, Regression Tree, Random Forest Regression and Gradient Boosting Regression Trees. The case study in the Beijing area shows that warehouse rent is closely related to its location and land price. Models considering multiple factors have better skill in estimating warehouse rent, compared to single-factor estimation. Additionally, tree models have better performance than the linear model, with the best model (Random Forest) achieving correlation coefficient of 0.57 in the test set. Deeper investigation of feature importance illustrates that distance from the city center plays the most important role in determining warehouse price in Beijing, followed by nearby real estate price and warehouse size.

**Keywords:** sharing warehousing, price estimation, machine learning.

## 1 Introduction

In today's logistics industry, insufficient capacity of self-owned warehouses and fluctuation in inventory have made sharing warehouses a useful option. However, more time and effort are needed for sharing warehouses to reach their full potential for supporting expanding supply chain businesses. In one 2015 survey [30], more than 75% of the participants were reported to suffer

significant inventory swings, with 31% of them faced with insufficient stock capacity and 26% with vacant storages. The paradox of both storage insufficiency and vacancy suggests a need for a smooth exchange of warehouse supply and demand information.

An ongoing digital marketplace transformation is taking place among various sections of the society. For instance, Airbnb, currently the largest online platform for people to lease or rent short-term lodgings, operates in 65,000 cities and had 31 billion of total valuation as of June 2017 [26]. Uber, the online driving service platform, had an average of 1 million rides of daily trips in 75 countries as of April 2017 [27].

Similar transformation happens in the shared warehousing as well. For example, Flexe, essentially the Airbnb of warehousing, provides a marketplace of spare storage space by gathering and releasing vacant warehouse information [29]. Craigslist, 58, and many other classified advertisements websites offer sections for people to post available warehouse messages.

In the platforms mentioned above, warehouse resources are usually archived into well-defined categories based on information offered by providers. Customers can make optimal decisions by filtering the listings with their demands, while platforms can offer customized integrated solutions when needed. This new business mode offers accessible information to both the supply and demand sides of the industry. It also encourages dedicated warehouses and self-hold vacant spaces to enter the shared warehouse market. A key question faced by the newcomers, as well as original warehouse providers, is how to price their warehouses in the open digital marketplace.

Theoretically, the optimal price is achieved from equilibrium in which the quantities of goods or services provided match the corresponding market's desire and ability to acquire the good or service. In an intransparent market, it's difficult, if not possible, for each participant to get the whole picture of the supply and demand information. This leads to unjustifiable pricing and inefficiency, which hurts the long-term prosperity of the market. Fortunately, such deficiency can be mitigated in the online marketplace, where information is more transparent to both sides of the transaction.

The availability of market information does not directly lead to optimal pricing. While an easy strategy is to take nearby shared warehouses as references to estimate a price, a more refined approach should estimate the price by considering an array of relevant factors, such as warehouse location, size, type, transportation, land price, etc. However, we lack a comprehensive empirical understanding of the pricing mechanism in this new emerging online marketplaces.

To bridge this gap, we apply the tool of Machine Learning to build warehouse rental pricing models. Such models can learn complex relations that are hidden within large amounts of real world data. They are especially helpful when our domain knowledge and understanding are limited. There are many previous works that use machine learning techniques for price estimation and prediction in various economic disciplines, such as stocks [9, 17], electricity [4, 23], online auctions [6, 18], real estate [10, 13, 14], lodging [1, 7, 11, 12, 22], high speed trains [24, 25], product lifecycles [19], and parking [16]. However, to our knowledge, this is the first work to gather real world data and analyse the pricing mechanism of the flourishing sharing warehouse market.

Our research develops a database consisting of real world sharing warehouse information from 58.com, one of the most popular classified advertisements website in China. The warehouse characteristics and its location information are selected as potential features in dominating the rental price. Four widely used machine learning algorithms, Linear Regression, Regression Tree, Random Forest Regression and Gradient Boosting Regression Trees, are applied to estimate warehouse pricing.

The rest of the paper is organized as follows. The data and materials are introduced in Section 2. We introduce the machine learning approaches in Section 3, followed by the results and discussion in Section 4. Finally, we summarize our conclusions.

Table 1: Shared warehouse data statistics

Attributes	Latitude(°)	Longitude(°)	Size( $m^2$ )	Rent (CNY/( $m^2$ day))
Mean	116.43	39.92	1,394.48	1.36
Std	0.16	0.14	2,311.61	1.26
Min	116.01	39.60	1.00	0.10
25%	116.31	39.84	101.75	0.60
50%	116.42	39.92	500.00	1.00
75%	116.56	40.00	1,500.00	1.50
Max	116.89	40.49	14,000.00	9.11

## 2 Data

Our major goal of this paper is to estimate sharing warehousing rental price using machine learning techniques. To explore this, we take the shared warehouse market in Beijing, one of the largest economic centers in China, as a case study. In this section, we describe the data set and relevant features for warehouse price estimation.

### 2.1 Data sets

We first create a dataset of sharing warehouses in the Beijing area. The data are acquired and organized based on warehouse rental posts on 58.com, one of the largest and most active general purpose classified websites in China. Posts during the period of November 2016 to January 2017 are archived using a web scraping framework in Python [21].

Fig. 1 gives an example of one rental listing. After removing incomplete and obviously erroneous data, in total we collected 2,462 rental listings, with each listing including information such as Title, District, Latitude, Longitude, Size and Rent across the Beijing area. Details on the data collection and cleaning processes can be found in Ma et al. [15]. The warehouse data statistics are shown in Table 1.

<b>Title: Rent Warehouse in Tongzhou</b>
<b>District: Tongzhou, Zhangjiawan</b>
<b>Location: Near East 6<sup>th</sup> Ring or Zhanjiawan Industrial Area</b>
<b>Type: Warehouse</b>
<b>Size: 1400 <math>m^2</math></b>
<b>Rent: 1000 yuan/month</b>

Figure 1: An example of Beijing Warehouse Rental Information Shown in Detailed Page of 58.com.

In addition to the warehouse pricing dataset, we also obtain 23,438 records of second-hand real estate rental information in the Beijing area in January 2017 from Lianjia.com, which is the website of one of the largest real estate agencies in China. Since a warehouse is a form of real estate, we assume that real estate rental prices are likely related to the shared warehouse rental price. The second hand real estate data statistics are described in Table 2.

Table 2: Second-hand Real Estate Data Statistics

Attributes	House Price (CNY/ $m^2$ )	Latitude( $^\circ$ )	Longitude( $^\circ$ )
Mean	59,542.34	116.44	39.94
Std	25,562.39	0.17	0.11
Min	3,949.00	115.9	39.62
25%	40,517.50	116.33	39.87
50%	54,299.50	116.41	39.94
75%	74,014.50	116.51	40.01
Max	149,963.00	116.95	40.47

## 2.2 Feature selection

To estimate the warehouse rental price, we extracted five features from the data sets, including Distance from City Center, Size, House Price (nearby second hand real estate price), Distance from the Closest House and District. A detailed explanation of these feature is given in the sequel.

We first consider the warehouse size and location, and present a general picture of their connection to rental price in Fig. 2. Each dot represents the location of a sharing warehouse, with the size indicating warehouse size and color indicating rental price. For the location-rent relationship, from Fig. 2, we can see that warehouses closer to the city center tend to have higher prices. For the size-rent relationship, we see that larger warehouses tend to have cheaper prices. Thus, we include Distance from City Center and Size as features in our warehouse rent estimation.

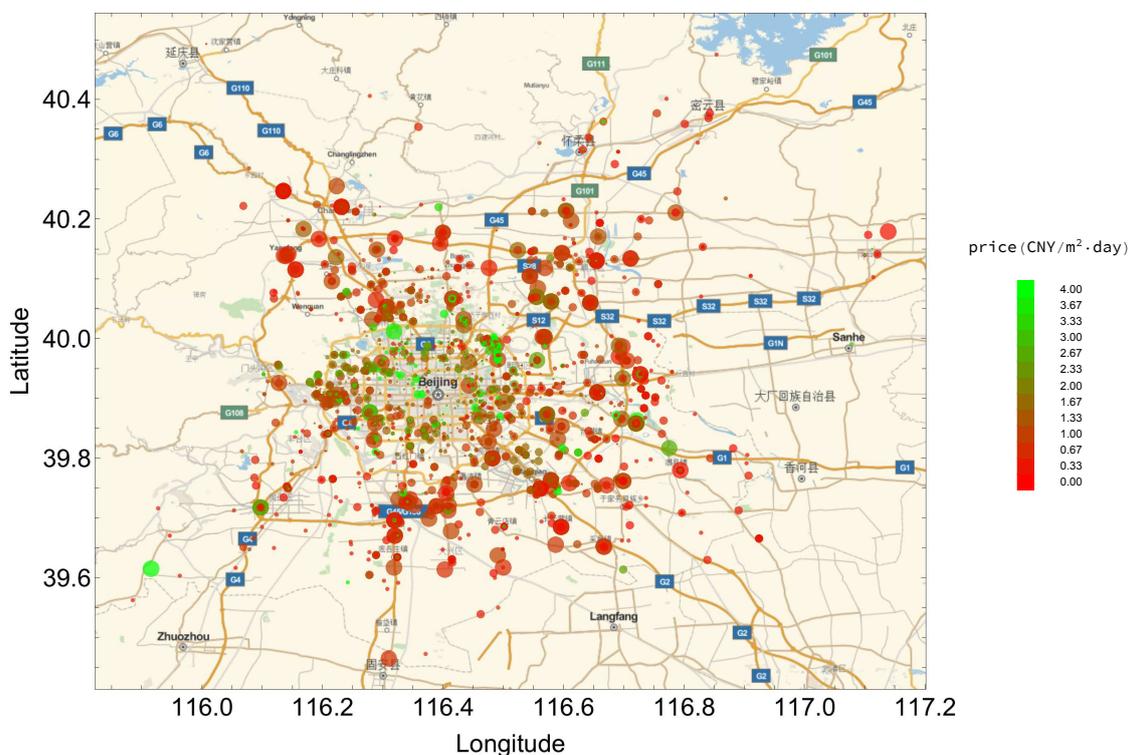


Figure 2: The Distribution of Shared Warehouse Listings in Beijing.

In addition, local real estate price can reflect the market value of the property. Since a warehouse is a kind of real estate, we assume there is a positive relationship between warehouse rent and house prices in the vicinity, with a stronger and more direct relationship when the two locations are closer. Fig. 3 shows the coordinates of warehouses and second-hand real estate. In this work, we computed the market price of the nearest second-hand house to the warehouse (House Price) and the distance between these two locations (Distance from the Closest House).

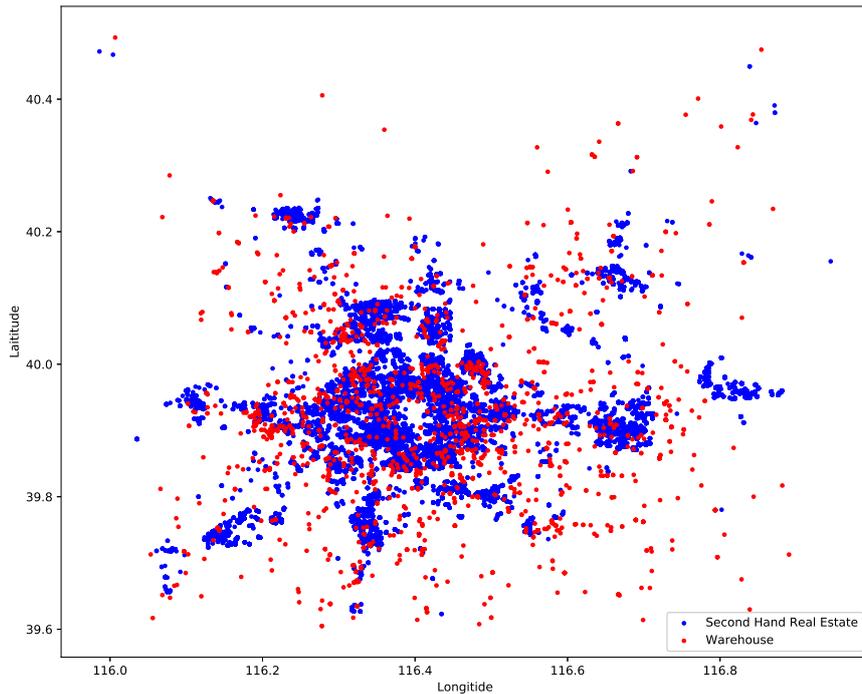


Figure 3: Coordinates of second hand real estate and shared warehouse in Beijing.

Finally, the location's District, a type of administrative division that is managed by local government, plays an important role in warehouse rental pricing. For example, as shown in Fig. 4, the warehouse rent is much more expensive in the Dongcheng and Xicheng districts than in other districts, and the lowest warehouse rent appears in the Miyun district. Thus, we select District as one of the features for estimating warehouse rent. Since it is a categorical feature, we convert it to a vector of fourteen binary indicator variables before modeling. A summary of the features is listed in Table 3, and an analysis of the relationship between several individual features and the warehouse rent is discussed in Section 4.1.

### 3 Methodology

In this section, we explain the machine learning techniques used in our experiment, which include:

- Linear Regression

Table 3: Feature statistics

Feature Name	Mean	Standard Deviation	Definition
Rent	1.36	1.26	Listed warehouse rent in online warehouse marketplace (Measured in CNY).
House Price	58,014.64	23,791.27	Listed price of nearest second hand house (Measured in CNY).
Distance from the Closest House	0.75	1.27	The distance between the location of a listed rental and the location of the nearest second hand house, computed as spherical distance with latitude and longitude (Measured in km).
Size	1,394.48	2,311.61	Listed warehouse size in online warehouse marketplace (Measured in m <sup>2</sup> ).
Distance from the City Center	17.17	10.35	The distance between the location of a listed rental and the city center, computed as spherical distance with latitude and longitude (Measured in km).
District Area 1	0.08	0.28	District Area: Changping. (Dummy variable)
District Area 2	0.23	0.42	District Area: Chaoyang. (Dummy variable)
District Area 3	0.13	0.33	District Area: Daxing. (Dummy variable)
District Area 4	0.01	0.11	District Area: Dongcheng. (Dummy variable)
District Area 5	0.02	0.15	District Area: Fangshan. (Dummy variable)
District Area 6	0.09	0.29	District Area: Fengtai. (Dummy variable)
District Area 7	0.15	0.36	District Area: Haidian. (Dummy variable)
District Area 8	0.00	0.06	District Area: Huairou. (Dummy variable)
District Area 9	0.00	0.04	District Area: Mengtougou. (Dummy variable)
District Area 10	0.00	0.05	District Area: Miyun. (Dummy variable)
District Area 11	0.04	0.21	District Area: Shijingshan. (Dummy variable)
District Area 12	0.06	0.23	District Area: Shunyi. (Dummy variable)
District Area 13	0.14	0.35	District Area: Tongzhou. (Dummy variable)
District Area 14	0.08	0.28	District Area: Xicheng. (Dummy variable)

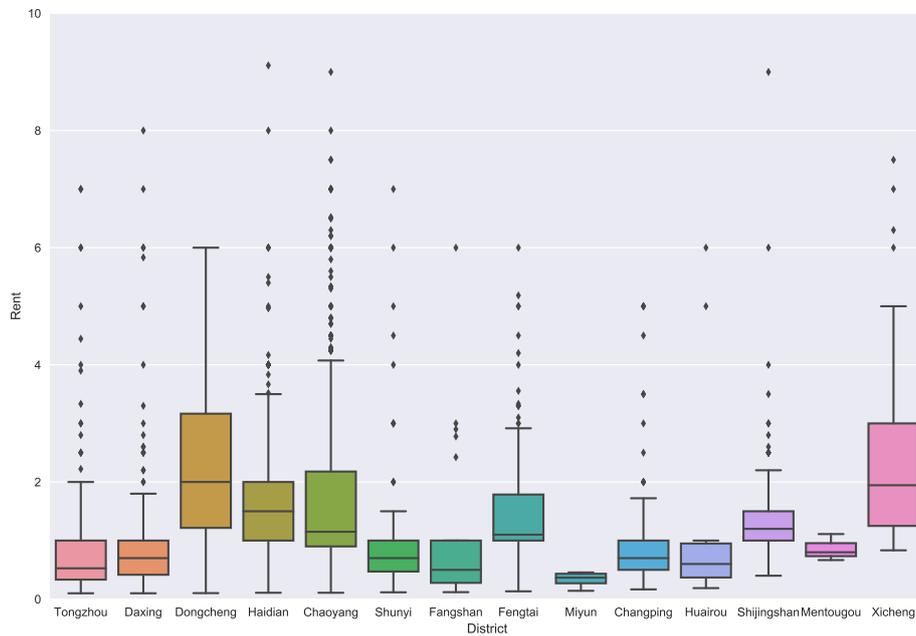


Figure 4: A boxplot showing the price of shared warehousing in Beijing, grouped by District.

- Regression Tree
- Random Forest Regression
- Gradient Boosting Regression Trees

A simple linear regression model constructs an optimized linear combination of the predictors (features). The relatively strict assumption of a linear relationship makes it difficult to fit on many real world problems, in which many factors may work together in a non-linear way to determine the predictand. The latter three models belongs to a broader family of tree-based learning algorithms. Tree algorithms are well suited for problems where features interact in nonlinear manners and no monad global model fits the predictor-predictand relation well. The basic idea behind these algorithms is to recursively partition the feature space until each cell of the partition is well modeled by a simple predictor (such as a constant value) [8]. Next we provide a more detailed explanation of the three tree-based algorithms.

### 3.1 Regression Tree

The regression tree, a variant of decision trees, is a popular method designed to approximate real-valued functions. Regression trees have a number of advantages. For one, they can be used without requiring feature pre-processing or normalization, since each feature is processed independently. Regression trees also tend to work well with data sets that have a mixture of continuous and categorical variables, and with features that are on very different scales, both of which are present in our research.

Basically, regression trees learn a series of explicit if/then rules based on feature values to predict the target value. Unlike linear regression, which holds a single predictive formula over

the entire feature space, regression trees break the global model into many local models using this recursive partitioning process, and then performs multiple (very simple) local regressions. There are two main parts to building a regression tree: growing the trees and pruning the trees.

The basic Regression Tree growing algorithm is as follows:

1. Start from a root node, containing all data points. Calculate the prediction for the leaf (typically, the mean value of the target values for the training points in that leaf), and the sum of squared errors for the leaf.
2. Search over all possible splits of all variables for the one which results in the smallest sum of square errors for its two leaves. If the sum of square errors is less than some user-defined threshold, or one of the resulting nodes contains less than a user-defined number of points, stop splitting. Otherwise, continue splitting data and creating new nodes.
3. For each new node, go to Step 1.

Once the tree's induction process is finished, pruning can be applied to improve the tree's generalization capacity by reducing its structural complexity. The number of cases in nodes can be used as the pruning criteria.

In the Regression Tree construction process, pre-pruning is used to avoid growing an overly complex tree. Allowing the tree to grow unpruned can result in overfitting, since the model may fit noise in the data. We control the maximum tree structural complexity using a max-depth parameter and a minimum number of samples per leaf. Max depth controls the maximum depth of the tree, limiting the total number of split points any decision can have. The min samples per leaf parameter is a threshold that controls how many data instances must be present in a leaf in order to consider splitting it further. To obtain robust and generalizable models in the experiment, many possible choices of these control parameters were assessed, ranging over depths from 1 to 50 and minimum sample sizes from 1 to 50. Via cross-validation, we found the best results at maximum depth 12 and minimum leaf samples set to 13.

### 3.2 Random Forest Regression

The Random Forest Regression algorithm, proposed by Breiman [3], is an ensemble method using decision trees as base learners. It combines many different individual trees into an ensemble, and introduces random variation while building each decision tree. The term "random" in random forest has a dual meaning: a subset of the data is randomly chosen to establish each tree, and features are randomly selected in each split test as well. The resulting ensemble of trees is averaged to produce an overall prediction, which reduces overfitting while allowing for complex individual learners.

Random Forests have several good characteristics, including high accuracy among current algorithms, efficiency on large data sets, the ability to handle high dimensional input variables and missing data. Thus, Random Forest Regression is widely used and often yields very good results on a variety of problems [20].

The Random Forest Regression algorithm is as follows:

1. Pick  $n_{tree}$  bootstrap samples from the dataset;
2. Build an unpruned regression tree based on each of bootstrap samples. When picking the best split for a node, a random subset of features is selected to be searched over, rather than finding the best split across all possible features;
3. Predict new data using the mean of  $n_{tree}$  trees' predictions.

Random Forest models can compute an internal unbiased estimate of the generalization error, called out-of-bag (or OOB). Given enough trees, the OOB estimate of error rate can be quite accurate.

The output of Random Forest Regression depends primarily on two parameters: maximum number of features and number of estimators. The maximum number of features parameter is the number of predictors chosen randomly at each tree node. It greatly influences the diversity of the random trees in the forest. When the parameter is low, the trees become more complex and diverse. However, if the parameter is high (e.g., close to the total number of features), the trees in the forest will tend to be very similar. A typical default setting of max features for regression is the log base two of the total number of features. The number of estimators parameter represents the number of trees in the ensemble. Ensembles reduce overfitting by averaging over more trees, but this increases computation. Increasing the number of trees tends out to be a better solution [2]. After testing, we chose the log base two of the total number of features (which is 4 in the experiment) and 3000 total trees.

### 3.3 Gradient Boosting Regression Tree

Gradient Boosting Regression Tree is another tree based ensemble method widely used in real world applications. It often yields excellent off-the-shelf accuracy on many problems. Unlike Random Forest Regression, which builds a forest of different trees in parallel, Gradient Boosting Regression Tree establishes a sequence of trees, each of which is intended to correct the mistakes of the previous trees in the series.

The basic idea of boosting is to convert weak learners to strong ones. Starting from a weak learner, we do iterations, where at each iteration, we change the weight distribution of the data and apply a new weak learner to the weighted data. This builds a sequence of different weak learners, and by combining those weak learners, we get a final, strong learner. When our model makes a large error (high cost) on a given data point, we assign a greater weight to that data point in next iteration, and assign less weight on the points for which the cost is less. If the cost of a weak learner is large, we give it a smaller contribution when combining the learners, and give greater emphasis on the learners for which the cost is less.

The Gradient Boosting Regression Tree algorithm for a loss or cost function  $L(y^{(i)}, f(x))$  is as follows [5]:

1. Initialize the model:

$$f_0(x) = \operatorname{argmin}_c \sum_{i=1}^M L(y^{(i)}, c)$$

2. Train K models iteratively  $k = 1, 2, \dots, M$ :

- (a) Compute the residual: for  $i = 1, 2, \dots, M$ :

$$r_{ki} = - \left[ \frac{\partial L(y^{(i)}, f(x^i))}{\partial f(x^i)} \right] f(x) = f_{k-1}(x)$$

- (b) Train a regressor for the residual and obtain the leaf nodes  $R_{kj}, j = 1, 2, \dots, J$
- (c) For  $j = 1$  to  $J$ , calculate:

$$c_{ki} = \operatorname{argmin}_c \sum_{x_i \in R_{kj}} L(y^i, f_{k-1}(x_i) + c)$$

(d) Update the model:

$$f_k(x) = f_{k-1}(x) + \sum_{j=1}^J c_{kj} I(x \in R_{kj})$$

3. Get the final strong learner:

$$f_s(x) = f_K(x) = \sum_{k=1}^K \sum_{j=1}^J c_{kj} I(x \in R_{kj})$$

In Gradient Boosting Tree Regression model, the number of estimators and the learning rate are important parameters. The first one controls the number of trees included in the ensemble. The second parameter works as a step size, scaling the contribution of each individual tree in reducing the loss. The effects of these two parameters interact: a smaller learning rate can often give better performance, but at the cost of requiring more trees in the ensemble (and thus more computation as well). Via cross-validation we find the best results by choosing ensemble size 4,000 with learning rate set to 0.01.

### 3.4 Model evaluation

For all the models, we take 70% of the data as a training set and use rest of the data as a test set. Two common score metrics are used for evaluating model performance: Root Mean Square Error (RMSE) and correlation coefficient ( $r$ ). The “best” model is the one with the lowest RMSE and the highest  $r$ .

$$RMSE = \frac{\sqrt{\sum_{i=1}^m (e_i - o_i)^2}}{m}$$

$$r = \frac{\sum (e_i - \bar{e})(o_i - \bar{o})}{\sqrt{\sum (e_i - \bar{e})^2 \sum (o_i - \bar{o})^2}}$$

where  $e$  is the model estimation,  $m$  is sample size, and  $o$  is the observation.

## 4 Results and discussion

In this section, we use each of the four modeling strategies to predict warehouse rental price. We demonstrate the performance of each model, and illustrate the importance of the various features.

### 4.1 Model performance

Before presenting the model performance, we show the correlation between rent and selected features in Fig.5. Rent and Distance from City Center are negatively correlated. Warehouses that are closer to the city center have higher rent variance. Considering the Size feature, 70% warehouses have sizes below  $1000m^2$ . In this range, size alone has little impact on Rent. For warehouses that are larger than  $5000m^2$ , rents are relatively small and stable. We also see that House Price is a good indicator for land price. Among all the features, House Price has the greatest impact on Rent ( $r = 0.34$ ). All three factors can affect the warehouse rent; however, the combination of them provides more information on predicting warehouse rental price. Below we illustrate model results considering all the features we described in Section 2.2.

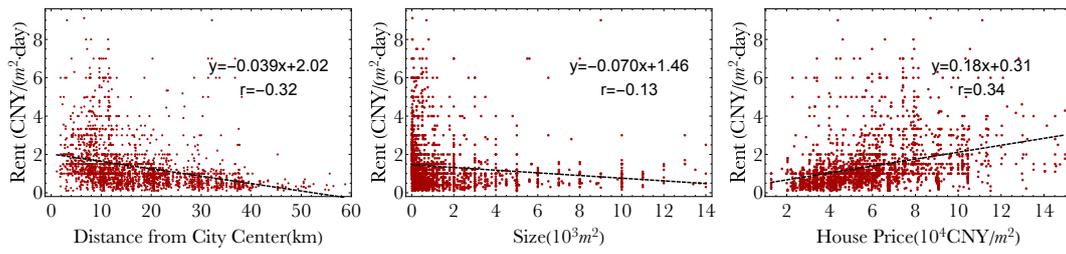


Figure 5: The relationship between rent and each selected feature. The fitted linear relation and its correlation coefficient are indicated in each sub-figure.

Fig. 6 shows the scatter plot of estimated and true warehouse rental prices of the four candidate models. Here we only show the test data performance. Generally, all models show a certain skill in estimating warehouse rental price. The tree based models have obviously higher skill compared to the Linear Regression Model. This reveals the nonlinear relationship between the factors and the rental price. Additionally, models for simulating lower rental price have better performance. When it comes to higher rental price, models tend to underestimate the rents.

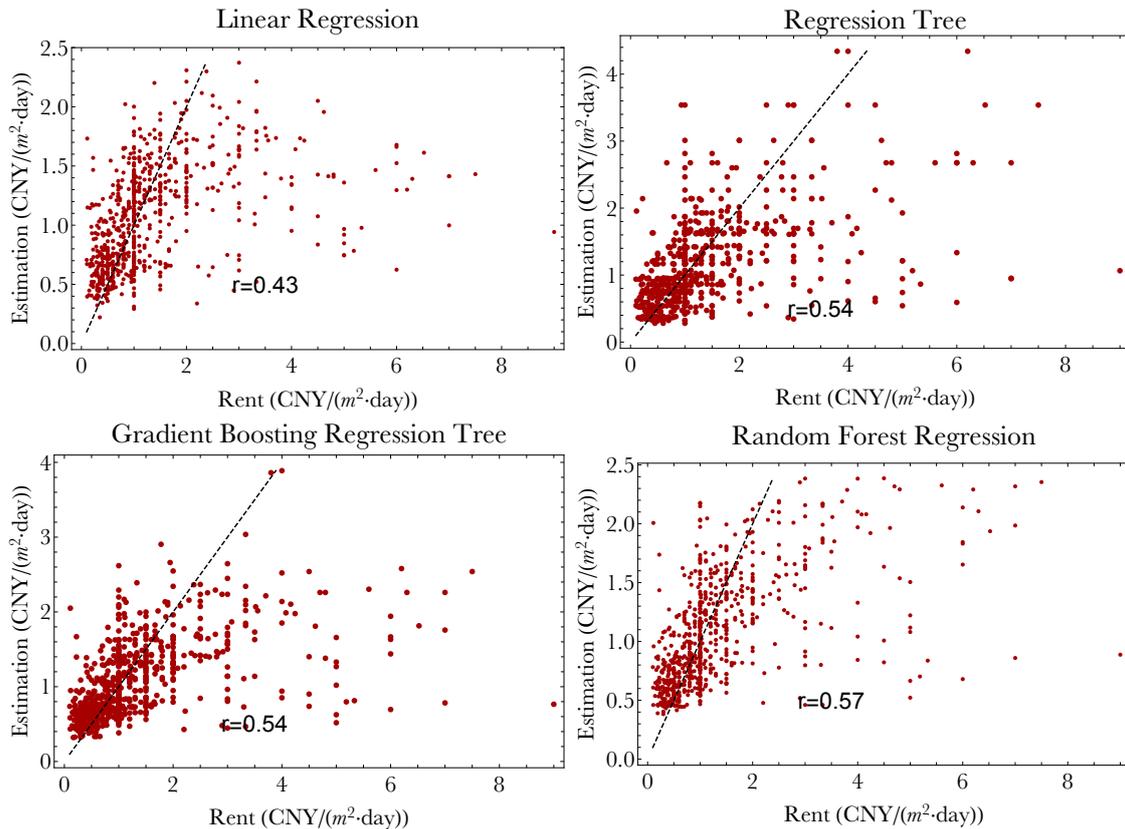


Figure 6: The relationship between the estimation and the target value for different models on the test set.

Table 4 shows results of RMSE and correlation coefficients for different models. Three tree based algorithms have smaller RMSE value and larger  $r$  value both on training set and test set, compared to Linear Regression. Within three tree models, the results of RMSE and  $r$  are similar. This indicates that the tree-based models are more capable of capturing data interconnection and estimating warehouse price, compared to Linear Regression. Of the four models, the Regression

Table 4: RMSE and correlation coefficient among different models

Model	Training Set		Test Set	
	RMSE (CNY/m <sup>2</sup> ·day)	r	RMSE (CNY/m <sup>2</sup> ·day)	r
Linear Regression	1.31	0.43	1.25	0.44
Regression Tree	1.02	0.63	1.05	0.54
Random Forest Regression	1.04	0.60	1.06	0.57
Gradient Boosting Regression Tree	1.05	0.68	1.07	0.54

Table 5: Error among different models on the test set

Model	Min	25%	Medium	75%	Max	Mean	Std
Linear Regression	-1.30	-0.27	-0.04	0.42	8.05	0.32	1.11
Regression Tree	-2.60	-0.28	0.03	0.38	7.93	0.22	1.03
Gradient Boosting Regression Tree	-1.94	-0.23	0.03	0.39	8.23	0.28	1.04
Random Forest Regression	-1.90	-0.24	0.00	0.40	8.11	0.28	1.03

Tree model has the smallest RMSE (1.05 CNY/m<sup>2</sup>·day), and Random Forest Regression has the highest correlation coefficient (0.57) on the test set.

Table 4 shows the error statistics for different models on test set. Linear Regression has the highest error mean and standard deviation and Regression Tree Model has smallest error mean and standard deviation.

To further analyze the error, we draw the error distributions of four models in different observation intervals on the test set in Fig.7. There is high relevance between the observed value and the model error. Large target values tend to have higher error mean and variance, no matter what model is used. These results could be related to data quality or the pricing mechanism. High prices may be boosted to aid subsequent bargaining, or list a fake price rather than representing the true warehouse value. Alternatively, warehouses with higher rent may be priced based on different criteria compared to lower rent warehouses.

## 4.2 Feature importance

The features predicted to be important in the model help us understand what features are driving the rent and what features are deemed important for each of the model. Feature importance is typically a number between 0 and 1 assigned to an individual feature. A feature importance of zero means the feature is not used at all in the prediction. A feature importance of one means the feature perfectly predicts the target.

For regression tree, the importance of a feature is simply the total reduction of the cost in sum of squares achieved by all splits based on that feature [3]. However, in random forest regression, there are two common measures of feature importance. The first is based on mean squared error (MSE), and the second is prediction accuracy of the out-of-bag portion of the data after permuting each feature. The difference between the two MSEs are then averaged over all trees and normalized by the standard error. The second measure is calculated on the training data used to grow the trees. Here we use the latter method as our measure of importance.

Table 6 shows the results of feature importance carried out by three tree algorithms. We can capture the consistency of feature importance in three models: Distance from City Center is the most influential feature, followed by Size and House Price. These features were chosen at an early level of the tree, comparing to the others. The District feature has the lowest importance value in all the models, which suggests that District has the smallest impact on Rent prediction.

The three models are slightly different in their feature importance order. Both Regression

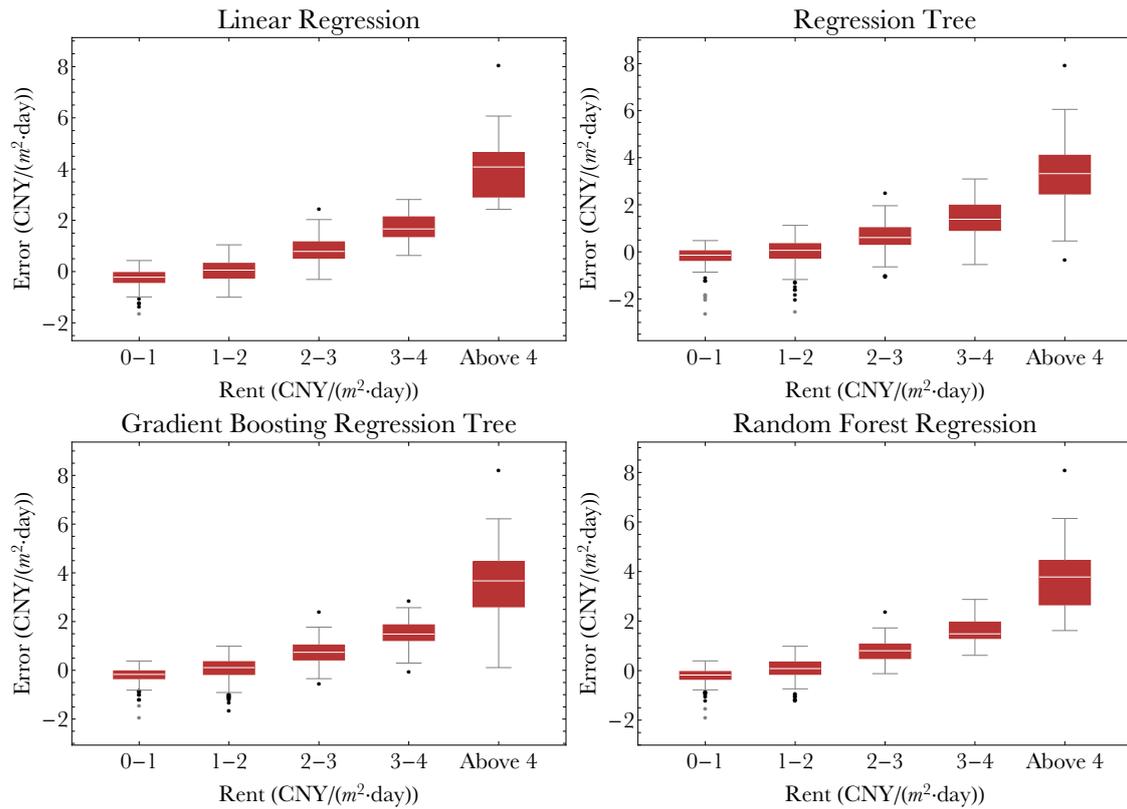


Figure 7: The relationship between error and target value of different models on the test set.

Tree and Random Forest Regression Model take House Price as the second most important feature, while it is Size in the Gradient Boosting Regression Tree.

## 5 Conclusion

The rental warehouse market is undergoing a transformation boosted by digital platforms. Both the supply and demand sides are faced with information transparency, which creates fluctuations in the pricing. To better understand the pricing mechanism in this evolving market and guide the pricing process for users, we collected and analyzed real-world rental warehouse data from the classified advertisement website.

We estimated the rental price of shared warehouses using five features selected from the data set: Distance from City Center, House Price, Distance from the Closest House, Size and District. Regression Tree, Random Forest Regression and Gradient Boosting Regression Tree along with Linear Regression were compared for predicting price. We used RMSE and correlation coefficient to provide model performance measurements. Results show that single factors exert considerable influence on rent pricing, while models considering multiple factors have more predictive skill. Tree based models provide better performance than Linear Regression, and the Random Forest method gave the best performance.

Feature importance analysis showed that the location, measured by distance from city center, plays the most important role in determining warehouse rental price, followed by local land price or warehouse size.

Estimation error is further analyzed conditioned on the warehouse rent. We found that high rent warehouses often correspond to higher error levels in the models. Thus, high rent warehouses

Table 6: Feature importance in tree based models

Feature Name	Importance (Regression Tree)	Importance (Random Forest Re- gression)	Importance (Gradient Boosting Re- gression Tree)
Distance from City Center	<b>0.434205</b>	<b>0.273986</b>	<b>0.287716</b>
Size	0.173837	0.158182	<b>0.217792</b>
House Price	<b>0.207980</b>	<b>0.233020</b>	<b>0.207745</b>
Distance from the Closest House	0.112518	0.164432	0.164071
District Area 1	0.002581	0.011886	0.007634
District Area 2	0.009909	0.026253	0.009557
District Area 3	0.005362	0.025396	0.013583
District Area 4	0.000000	0.000730	0.009991
District Area 5	0.011750	0.009544	0.021440
District Area 6	0.000218	0.007067	0.003513
District Area 7	0.011849	0.025661	0.019505
District Area 8	0.003037	0.000000	0.000460
District Area 9	0.000000	0.000000	0.000000
District Area 10	0.000000	0.000000	0.000000
District Area 11	0.016273	0.011826	0.020358
District Area 12	0.002113	0.002895	0.000325
District Area 13	0.006511	0.046599	0.009163
District Area 14	0.001858	0.002523	0.007147

may require more factors to be considered in estimating their price. However, the high prices might also be due to price boosting or fake posts.

Despite the various factors and models used here, there are still considerable errors in estimating the warehouse rents. On the one hand, more features from the real world should be considered, such as text information that reveal social or psychological factors influencing warehouse pricing. On the other hand, it may also be true that in the growing free market, not all prices are justified; the data are intrinsically noisy. Our methodology here provides an attempt to extract useful information from the complicated real world market.

## Acknowledgment

This work was supported by the National Natural Science Foundation of China(Grant No.71390334), Beijing Program of Philosophical and Social Science (Grant No.B16HZ00060) and Junior Fellowship for CAST Advanced Innovation Think-tank Program (Grant No.B16M00150).

## Bibliography

- [1] Antipov, E. A.; Elena, B. P. (2012); Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics, *Expert Systems with Applications*, 39(2), 1772–1778, 2012.
- [2] Breiman, L. (1996); Bagging predictors, *Machine Learning*, 24(2), 123–140, 1996.
- [3] Breiman, L. (2001); Random forests, *Machine Learning*, 45(1), 5–32, 2001.

- 
- [4] Chen, X.; Dong, Z. Y.; Meng, K.; Xu, Y.; Wong, K. P.; Ngan, H. W. (2012); Electricity price forecasting with extreme learning machine and bootstrapping, *IEEE Transactions on Power Systems*, 27(4), 2055–2062, 2012.
- [5] Elith, J.; Leathwick, J. R.; Hastie T. (2008); A working guide to boosted regression trees, *Journal of Animal Ecology*, 77(4), 802–813, 2008.
- [6] Ghani, R. (2005); Price prediction and insurance for online auctions, *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, 411–418, 2005.
- [7] Gutt, D.; Herrmann, P. (2015); Sharing means caring? Hosts’ price reaction to rating visibility *ECIS*, 2015.
- [8] Hastie, T.; Tibshirani, R.; Friedman, J. (2001); *The Elements of Statistical Learning*. Springer, 2001.
- [9] Kim, K. J. (2003); Financial time series forecasting using support vector machines, *Neurocomputing* 55(1), 307–319, 2003.
- [10] Kusan, H.; Osman A.; Ilker O. (2010); The use of fuzzy logic in predicting house selling price, *Expert systems with Applications*, 37(3), 1808–1813, 2010.
- [11] Li, J.; Moreno, A.; Zhang, D. J. (2015); Agent behavior in the sharing economy: Evidence from Airbnb, *Ross School of Business Working Paper Series*, 1298, 2015.
- [12] Li, Y.; Wang, S.; Yang, Pan, Q.; Tang, J. (2017); Price recommendation on vacation rental websites, *Proceedings of the 2017 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics*, 399–407, 2017.
- [13] Limsombunchai, V. (2004); House price prediction: hedonic price model vs. artificial neural network, *New Zealand Agricultural and Resource Economics Society Conference*, 25–26, 2004.
- [14] Liu, J. G; Zhang, X. L.; Wu, W. P. (2006); Application of fuzzy neural network for real estate prediction, *International Symposium on Neural Networks*, 1187–1191, 2006.
- [15] Ma, Y. X.; Zhang, Z. J.; Pan, B. X. (2017); Reveal status quo of Beijing warehouse in open market, *Logistics, Informatics and Service Sciences (LISS), 2017 International Conference on. IEEE*, 2011–2017, 2017.
- [16] Maric, M.; Gracanin, D.; Zogovic, N.; Ruskic, N.; Ivanovic, B. (2017); Parking search optimization in urban area, *International Journal of Simulation Modelling*, 16(2), 2017.
- [17] Patel, J.; Shah, S.; Thakkar, P.; Kotecha, K. (2015); Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques, *Expert Systems with Applications*, 42(1), 259–268, 2015.
- [18] Raykhel, I.; Ventura, D. (2008); Real-time automatic price prediction for eBay online trading, *IAAI*, 2009.
- [19] Rehar, T.; Ogrizek, B.; Leber, M.; Pisnik, A.; Buchmeister, B. (2017); Product lifecycle forecasting using system’s indicators, *International Journal of Simulation Modelling*, 16(1), 2017.

- [20] Segal, M. R. (2004); Machine learning benchmarks and random forest regression, *Center for Bioinformatics Molecular Biostatistics*, 2004.
- [21] Scrapy Community. (2015); *Scrapy: A Fast and Powerful Scraping and Web Crawling Framework*, [http:// scrapy.org/doc/](http://scrapy.org/doc/).
- [22] Wang, D.; Nicolau, J. L. (2017); Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb. com, *International Journal of Hospitality Management*, 62, 120–131, 2017.
- [23] Yamin, H. Y.; Shahidehpour, S. M.; Li, Z. Y. (2004); Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets, *International journal of electrical power & energy systems*, 26(8), 571–581, 2004.
- [24] Zhang, D. (2017); High-speed Train Control System Big Data Analysis Based on Fuzzy RDF Model and Uncertain Reasoning, *International Journal of Computers, Communications & Control*, 12(4), 2017.
- [25] Zhang, D.; Sui, J.; Gong, Y. (2017); Large scale software test data generation based on collective constraint and weighted combination method, *Technical Gazette*, 24(4), 1041–1049, 2017.
- [26] [Online]. Available: [www.expandedramblings.com/index.php/airbnb-statistics/](http://www.expandedramblings.com/index.php/airbnb-statistics/), Accessed on 10 September 2017.
- [27] [Online]. Available: [www.expandedramblings.com/index.php/uber-statistics/](http://www.expandedramblings.com/index.php/uber-statistics/), Accessed on 10 September 2017.
- [28] [Online]. Available: [www.dhl.com/](http://www.dhl.com/), Accessed on 10 September 2017.
- [29] [Online]. Available: [www.flexe.com/](http://www.flexe.com/), Accessed on 10 September 2017.
- [30] [Online]. Available: [www.flexe.com/Flexe-capacity-eco/](http://www.flexe.com/Flexe-capacity-eco/), Accessed on 10 September 2017.

# Control-Scheduling Codesign for NCS based Fuzzy Systems

P.E. Mendez-Monroy, I. Sanchez Dominguez, A. Bassam, O. May Tzuc

**Paul Erick Mendez-Monroy\***, **Israel Sanchez Dominguez**

IIMAS-Merida Universidad Nacional Autonoma de Mexico,  
Parque Cientifico y Tecnologico de Yucatan,  
5.5Km Carretera Sierra Papacal - Chuburna,  
C.P. 97302 Sierra Papacal, Yucatan, Mexico

\*Corresponding author: erick.mendez@iimas.unam.mx  
israel.sanchez@iimas.unam.mx

**Ali Bassam, Oscar May Tzuc**

Facultad de Ingenieria, Universidad Autonoma de Yucatan,  
Av. Industrias no Contaminantes,  
Apdo. Postal 150 Merida, Yucatan, Mexico  
baali@correo.uady.mx, maytzuc@gmail.com

**Abstract:** In the present paper, a fuzzy codesign approach is proposed to deal with the controller and scheduler design for a networked control system which is physically distributed with a shared communication network. The proposed fuzzy controller is applied to generate the control with different sampling-actuation periods, the configuration supposes a strict actuation period disappears the jitter. The proposed fuzzy scheduling is designed to select the sampling-actuation period. So, the fuzzy codesign reduces the rate of transmission when the system is stable through the scheduler while the controller adjusts the control signal. The fuzzy codesign guarantees the stability of all the system if the network uncertainties do not exceed an upper bound and is a low computational cost method implemented with an embedded system. An unstable, nonlinear system is used to evaluate the proposed approach and compared to a hybrid control, the results show greater robustness to multiple lost packets and time delays much larger than the sampling period. This paper is an extension of [20].<sup>a</sup>

**Keywords:** codesign, dynamic scheduling, fuzzy control, networked control system.

---

<sup>a</sup>Reprinted (partial) and extended, with permission based on License Number 4275590998661 IEEE, from "Electrical Engineering, Computing Science and Automatic Control, 2017 14th International Conference on".

## 1 Introduction

This paper is an extension of [20]. The system model with sampling-actuation periods in [20] is modified using the one-step control input, the network imperfection estimation is improved, the feedback matrices are calculated in a simpler way. Finally, the analysis of stability and the analytical codesign are also presented.

Networked control systems (NCSs) are composed of physically distributed agents that can sense the environment, act on it, and communicate with one through a communication network to achieve some common goals. These characteristics have made them a topic of current interest in the control area.

By including a communication network within the control loop, considerations are presented for its design. Among these considerations the most important are time delays [13] [27] [11] [12], packet losses [13] [27] [11] [12], signal quantization [14] [21] and scheduling [6] [28]. These have been investigated with results reported in the literature. In addition, because of the advantages of reduced system wiring, simple installation, increased system flexibility and resources sharing,

NCSs have been finding applications in DC motors [5] [9], robot control [28], vehicle robot [25] and ball maglev system [11], among others.

Time delays in NCSs are the major cause of system performance deterioration and potential system instability. Time delays have been modelled by using various formulations such as constant delay [22], independently random delay [1] and random delay governed by Markov chain [32]. Sometimes, the time delays are included as time-varying input delays of the system [30]. Yi, An and Choi performed a control system over a wireless network with communications only in the control channel. They employ a full-order observer to estimate the system states without time delays. The control-actuator time delay is measured at the plant and the observed states are used as a predictor to generate the system state with the expected time delay. Finally, they employ a LQR control to generate the control signal based on the next estimated states.

Therefore, the analysis and synthesis of NCSs with both time delays and packet losses is a persistent problem in the challenging but practical problem. In the literature, some important methodologies, such as stochastic control [11], predictive control [5] [10] [30], robust control [12], and state feedback control [31], are proposed to compensate time delays and/or packet losses have been proposed.

At the last years, fuzzy logic control has received great attention from academic and industrial communities. More recently, the fuzzy control has developed strategies for NCS, In the work of Peng and Yang [24], a delay distribution-dependent design method for NCS Takagi Sugeno fuzzy systems [26] was proposed taking into consideration of the probabilistic interval distribution of the communication delay. Tong, Qian and Lui [29] used a fuzzy predictive controller to counteract time delays in the feedback channel. Where the fuzzy controller estimates the variations of the control signal based on the differences between the reference error and the control error applied. Chai et al. [4] investigated the state feedback and dynamic output feedback controller design for membership functions and time delays in premise variables into the controller design. The resulting conditions were expressed in terms of SOS-based inequalities.

In the case of lost packets, commonly the effect has been modelled by a Bernoulli process and strategies in static/dynamic output feedback and model predictive control problems for discrete-time T-S systems with lost packets were investigated in [7] [34]. In [15], Li, Wu, and Feng used a fuzzy model to describe a nonlinear plant with an output feedback controller  $H_\infty$  and modelling the lost packets as a Bernoulli random binary distribution.

It is noticed that most of the existing control methodologies for NCSs adopt a sampling period regardless of network Quality-of-Services (QoS) variations. In practical circumstances, the network QoS always fluctuates due to changes in the traffic load and available network resources.

In regard to QoS, Tipsuwan and Chow [28] proposed a gain scheduling controller for NCSs, where the control parameters were adjusted on-line based on network QoS variations and Chow [6] optimized the control parameters for gain scheduling controller to improve the NCSs performance. However, these works only focused on the controller design. More recently, Benítez et al. [3] presented a frequency control of multiple network control systems, this takes into account information from the network transmissions, where asymptotic stability of the systems is ensured when the time delay is bounded.

This paper shows a fuzzy NCS codesign controller-scheduler to adapt simultaneously the control signal and sampling period with estimated network imperfections. This introduces a neural model to estimate the time delay and lost packets as a compound time. A fuzzy control with the estimated compound time as the antecedent part is used to minimize the network effects. Finally, a fuzzy scheduler is designed to modify the sampling-actuation period based on the system performance and network utilization.

The paper is organized as follows: section 2 introduces a dynamic model based on actuation periods and shows a recurrent neural network to estimate some network imperfections. In section 3 the fuzzy model with different actuation periods and the stability analysis is presented. Section 4 summarizes the design of a fuzzy scheduler to modify the actuation period. Section 5 presents an analytical codesign and section 6 presents the experimental case to evaluate the codesign performance, it is compared with a hybrid controller. Finally, the conclusions are provided in section 7.

## 2 Preliminaries

### 2.1 Periodic Actuation Model

NCS is defined as: A feedback control system closed via a communication channel, it may be shared with other control loops or nodes outside the control system. A spatially distributed system for a single control loop is shown in Figure 1, where the sensor, controller, and actuator nodes exchange information via a communication network.

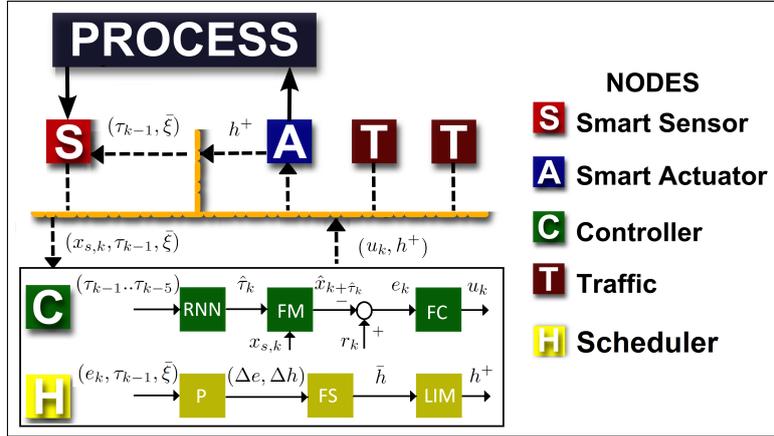


Figure 1: Configuration of NCS with traffic nodes

The most general continuous state-space representation of a linear system with  $m$  inputs,  $p$  outputs and  $n$  state variables is written in the following form

$$\begin{aligned} \dot{x}(t) &= Ax + Bu(t) \\ y(t) &= Cx(t) \end{aligned} \quad (1)$$

where  $A \in \mathbb{R}^{n \times n}$ ,  $B \in \mathbb{R}^{n \times m}$  and  $C \in \mathbb{R}^{p \times n}$  are the system, input and output matrices of the continuous-time state space respectively.  $x_k \in \mathbb{R}^n$  is the process state vector,  $u_k \in \mathbb{R}^m$  and  $y_k \in \mathbb{R}^p$  are the inputs and outputs of the process.

The continuous linear system Eq. 1 can be discretized assuming a zero-order hold [2] for the input vector with a sampling period  $h$  to

$$\begin{aligned} x_{k+1} &= \Phi_h x_k + \Gamma_h u_k \\ y_k &= Cx_k \end{aligned} \quad (2)$$

where the matrices  $\Phi_h \in \mathbb{R}^{n \times n}$  and  $\Gamma_h \in \mathbb{R}^{n \times m}$  are obtained by

$$\Phi_h = e^{Ah}, \quad \Gamma_h = \int_0^h e^{As} B ds \quad (3)$$

For standard closed-loop operation of the discrete system (Eq. 2), a controller can be designed using feedback control as follows

$$u_k = Kx_k, \quad K \in \mathbb{R}^{m \times n} \quad (4)$$

where  $K$  is the state feedback matrix obtained using standard control design methods.

The application of the control signal (Eq. 4) to the process forces to computing it with *zero* time. Nevertheless, this is physically impossible even for processor-based systems due to algorithm computational time.

Taking into account this limitation. The discrete model (Eq. 2) can be augmented to cope with a time delay due to the insertion of a network/processor within a control loop, as in the case of NCS [2]. The standard model that incorporates a time delay  $\tau$  less than one sampling period ( $\tau < h$ ), is

$$x_{k+1} = \Phi_h x_k + \Phi_{h-\tau} \Gamma_\tau u_{k-1} + \Gamma_{h-\tau} u_k \quad (5)$$

The equation 5 has been often taken as the essential control model for design and analysis of NCS. This model assumes a time reference given by the sampling instants with a fixed time delay from sampling to actuation. However, this model is useless if the time delay is variable and/or greater than one sampling period or the sampling interval is variable. [17].

The task execution model proposed is shown in figure 2. It aims is to use strict periodic sampling and actuation [18] into the space state model decreasing the variability in the time delays and the sampling intervals. This model estimates the states as a function of the actuation periods [16], making only necessary to estimate the compound time in a set of multiples of the actuation period.

The model synchronizes the operation of each control loop at the actuation instants. Hence,  $t_k$  is the actuation instant, the actuation interval is the time elapsed between consecutive actuation instants, named  $t_{k-1}$  and  $t_k$ ,  $h$  is the actuation period. Within this actuation interval, the system state is sampled, named  $x_{s,k}(t_{s,k}) \in [t_{k-1}, t_k]$  where  $t_{s,k}$  is the sampling time recorded. Eq. 6 represents a time delay  $\tau_k$  used to estimate the state at the actuation instant and the Eq. 7 represents the discrete system with periodic actuation.

$$\tau_k = t_k - t_{s,k} \quad (6)$$

$$\hat{x}_k = \Phi_{\tau_k} x_{s,k} + \Gamma_{\tau_k} u_{k-1} \quad (7)$$

Finally, making use of  $\hat{x}_k$ , the control command is computed as

$$u_k = K\hat{x}_k, \quad K \in \mathbb{R}^{m \times n} \quad (8)$$

where  $K$  is the feedback matrix that is designed in next section.

The control command  $u_k$  is held constant within actuation period with a zero-order Hold (ZOH).

At each control cycle the information flow in the NCS, the sensor node begins sampling the process  $x_{s,k}$  in time  $t_{s,k}$ , the time delay and lost packets are estimated. It sends the data to the codesign node where is used to generate  $u_k$  (Eq. 6)-(Eq. 8), also the scheduler generates the next actuation period if it is necessary. The control and period are sent to the actuator node applying the control to the process. Finally, the actuator node sends  $t_k$  and the period to the sensor node to apply the sampling period and calculate the time delay and lost packets. The cycle starts again.

With the strict periodic sampling and actuation  $h$ , the time delay  $\tau$  is restricted to multiples of the actuation period and the sampling intervals can be used to control the network bandwidth consumption. The model has several properties for controllers, it is compatible with standard

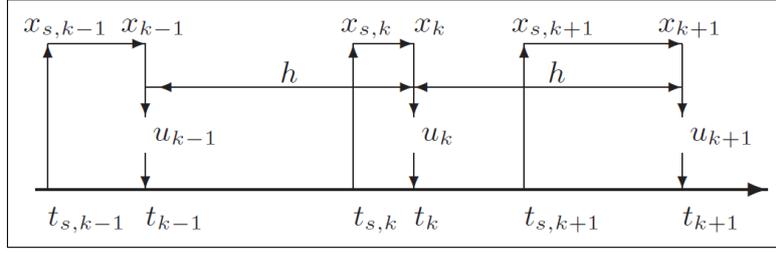


Figure 2: Periodic Actuation Model

scheduling because does not demand any specific timing constraints. The scheduling jitter is absorbed with a priori knew of time reference  $t_k$  and the sampling jitter disappears using the periodic actuation model (Eq. 7) absorbs the irregular sampling and variable actuation intervals.

The next section presents an approach to estimate the time delay and lost packets with a recurrent neural network.

## 2.2 Time delay and lost packets estimation

One major challenge for NCS design is the effect of time delays and lost packets in a control loop. The time delays occur when the system components exchange data across the network. It can degrade the performance or even destabilize the system. The time delay  $\tau_k$  assumes lower and upper bounds. In case of time delays is

$$0 < \tau_{min} \leq \tau_k \leq \tau_{max} \quad (9)$$

On the other hand, lost packets can be the consequence of a link failure, generated purposefully to avoid congestion or guarantee the most recent data to be sent. Normally, feedback controllers can tolerate a certain amount of lost packets. However, consecutive lost packets have an impact of degradation on the overall system performance. Hence, the next actuation period is a compound time  $v$  between time delays and lost packets formed as follows

$$v \equiv t_{k+1} - t_k = (\bar{\xi} + 1)h + \tau_{k+1} - \tau_k \quad (10)$$

where  $h$  denotes the actuation period,  $t_k$  the actuation instant and  $\bar{\xi}$  the estimated lost packets.

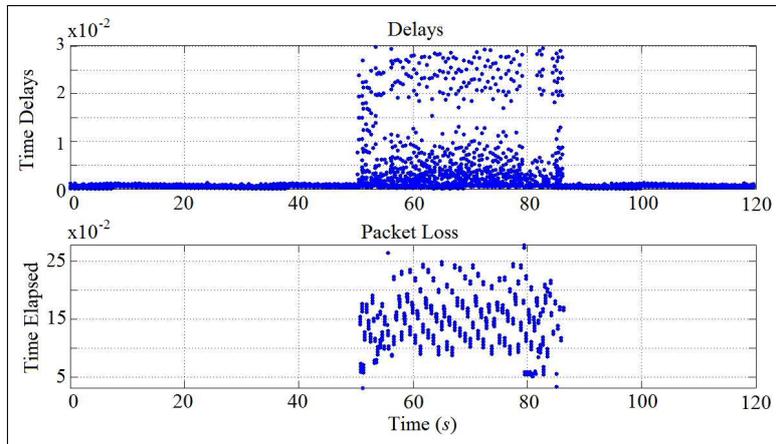


Figure 3: Time delays and time elapsed between lost packets with traffic (50 - 80 s).

Eq. 10 defines a compound time with the analytic bounds for time delay and lost packets, the aim is to design a recurrent neural network (RNN) using compound time values as input, and

time delay forecast as output. The collection of time delay data ( $\tau_k$ ) is analyzed and modelled to achieve the lowest discrepancies between the observed delay and the predicted delay (Fig. 3). Figure 3 illustrates the time delays and time elapsed between consecutive lost packets. Heavy traffic is generated into 50-80 s showing a maximum compound time of 300 ms.

The RNN is formed with three layers (Fig. 4), the input layer with 6 delayed inputs, one feedback input and bias, the hidden layer with 10 tansig nonlinear neurons, and the output layer with one linear discriminatory purelin neuron and bias.

$$\begin{aligned}
 I &= \{\tau_{k-i}, \hat{O}_{k-1}, b_1\} & i &= 1 - 6 \\
 H_j &= \text{tansig} \left[ \sum_{i=1}^7 IW_{ij} I_i \right] \\
 \hat{\tau}_{k+1} = O &= \text{purelin} \left[ \sum_{j=1}^{10} OW_j H_j + b_2 \right]
 \end{aligned}
 \tag{11}$$

The Levenberg-Marquardt algorithm was used for training. The number of hidden nodes selected was the RNN with the best validation performance.

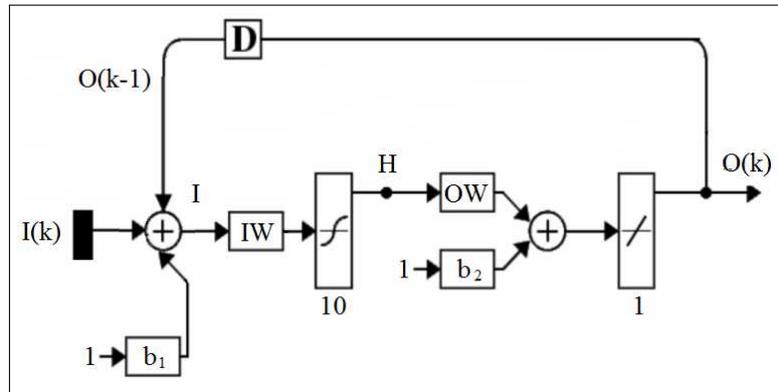


Figure 4: Structure of recurrent neural Network for time delay estimation

### 3 Fuzzy control

A new modelling method for nonlinear NCS with compound time is presented in this section. The process under consideration is a nonlinear discrete-time system represented by the TSK fuzzy model [26], It has the compound time  $\hat{v}$  as antecedent input and linear discrete models (Eq. 7) with different sampling periods  $h_j$  as consequent output. By modelling the system dynamics in function of the compound time.

So, defining  $r$  fuzzy rules, the  $j$ -th rule has form

$$\text{if } \hat{v} \text{ is } \alpha_j \text{ then } x_j = \Phi_j x_{s,k} + \Gamma_j u_{k-1}
 \tag{12}$$

where  $x_k \in \mathbb{R}^n$  is state vector,  $u_k \in \mathbb{R}^m$  is input vector,  $\alpha_j$  is the  $j$ -th membership function.

The overall fuzzy model is:

$$\hat{x}_k = \sum_{j=1}^r \psi_j [\Phi_j x_{s,k} + \Gamma_j u_{k-1}]
 \tag{13}$$

with the normalized fire strength  $\psi_j$  as

$$\sum_{j=1}^r \psi_j = 1 \quad \psi_j \geq 0 \quad \psi_j = \frac{\alpha_j}{\sum_{s=1}^r \alpha_s}
 \tag{14}$$

and

$$\alpha_j = \exp \left[ -\frac{[\hat{v} - \rho_j]^2}{\sigma_j^2} \right] \quad (15)$$

$\alpha_j$  is a Gaussian membership function with parameters  $[\rho_j, \sigma_j]$ . On the other hand,  $[\Phi_j, \Gamma_j]$  are the matrices of  $j$ -th linear discrete model discretized with a sampling period  $[h_j, j = 1 \dots r]$ , the discrete local models are:

$$x_j = e^{[Ah_j]}x_{s,k} + \int_0^{h_j} e^{As} ds B u_k = \Phi_j x_{s,k} + \Gamma_j u_{k-1} \quad (16)$$

So,  $[h_j, \rho_j, \sigma_j]$  for  $j = 1, \dots, r$  are assigned by user according to range of the compound time.

With this fuzzy model, the estimated system state is obtained by compensating the time delays, variable sampling intervals and lost packets. The action is to smoothly switch between discrete models to generate the best estimate of the state according to the estimated compound time  $\hat{v}$ .

Once designed the fuzzy model  $\hat{x}_k$  using the estimated compound time  $\hat{v}$  a fuzzy controller is proposed. This is a fuzzy feedback control law like:

$$u_k = - \sum_{j=1}^r \psi_j K_j \hat{x}_k \quad j = 1, \dots, r \quad (17)$$

where  $K_j$  is the feedback matrix of the  $j$ -th fuzzy rule. This control law is designed like a LQR (Linear Quadratic Regulator) [33] to minimize a performance index. The control design by LQR for each local model requires the algebraic solution of the Ricatti equation for the  $H_j$  matrix. So, the feedback matrices are calculated like:

$$K_j = R_j^{-1} \Gamma_j^T H_j \quad j = 1, \dots, r \quad (18)$$

The closed loop system is:

$$x_{k+1} = \sum_{i=1}^r \sum_{j=1}^r \alpha_i \beta_j [\Psi_i - \Gamma_i K_j] x_k = \sum_{i=1}^r \sum_{j=1}^r \alpha_i \beta_j \Lambda_{ij} x_k \quad (19)$$

with  $\Lambda_{ij} = \Phi_i - \Gamma_i K_j \quad i = 1, \dots, r \quad j = 1, \dots, r$

The properties of the antecedent part (Eq. 14) are considered for the stability analysis of fuzzy control  $u_k$ , the Eq. 20 are complementary properties of fuzzy sets.

$$\begin{aligned} \psi_i \psi_j &\geq 0 \\ \sum_{i=1}^r \sum_{j=1}^r \psi_i \psi_j &= 1 \\ \sum_{i=1}^r \psi_i^2 + 2 \sum_{i,j}^{i < j} \psi_i \psi_j &= 1 \end{aligned} \quad (20)$$

### 3.1 Stability analysis

Based on the properties of fuzzy control and assume that two-overlapped fuzzy memberships at most, a stability analysis of closed loop fuzzy control is presented. First, it is necessary to define the following lemma to prove stability analysis.

**Lemma 1.** [8] For any real matrices  $A_i, B_i$  for  $1 \leq i \leq r, P > 0 \in \mathbb{R}^{n \times n}$ , we have

$$2 \sum_i^r \sum_j^r \psi^i \psi^j A_i^T P B_j \leq \sum_i^r \sum_j^r \psi^i [A_i^T P A_i + B_i^T P B_i] \quad (21)$$

where the normalized fire strength  $\psi^i$  has the properties showed in Eq. 20 and Eq. 14.

So, the stability of the fuzzy model (13) and the fuzzy control (17) can be guaranteed if the following Theorem is fulfilled.

*Theorem 1.* The equilibrium state  $x_e = 0$  of closed loop system with control input (19 with two-overlapped fuzzy memberships at most, is asymptotically stable in the large, if there exist  $\mu$  positive definite matrices  $P_s = P_s^T > 0$  such that:

$$[\Lambda_{ij} + \Lambda_{ji}]^T P_s [\Lambda_{ij} + \Lambda_{ji}] - 2P_s < 0 \quad i \in S_s \quad s = 1, \dots, \mu \tag{22}$$

$$\Lambda_{ii}^T P_s \Lambda_{ii} - P_s < 0 \quad i \in S_s \quad j \in S_s \quad i < j \in S_s \tag{23}$$

$\Lambda_{ij} = \Phi^i - \Gamma^i K^j$  where  $S = \{S_1, S_2, \dots, S_\mu\}$  are  $\mu$  regions where two fuzzy rules are fired at most (two overlapped fuzzy memberships), where  $S_s$  contains the indexes of fired membership functions in  $\mathbf{s}$  region.

**Proof:** We suppose that there exist  $\mu$  matrices  $P_s = P_s^T > 0$  so (22) and (23) are satisfied. Considering a candidate Lyapunov function like:

$$V_k = \sum_{s=1}^{\mu} \lambda_s [x_k^T P_s x_k] \tag{24}$$

where

$$\lambda_s [\hat{v}] = \begin{cases} 1 & \hat{v} \in S_s \\ 0 & \hat{v} \notin S_s \end{cases} \quad \sum_{s=1}^{\mu} \lambda_s [\hat{r}] = 1 \tag{25}$$

It can be easily showed that  $V[0] = 0$ ,  $V_k > 0$  for  $x_k \neq 0$ , and  $V[x] \rightarrow \infty$  as  $\|x_k\| \rightarrow \infty$ , it is only sufficient shows that  $\Delta V[x_k] < 0$  to prove that  $V_k$  is a Lyapunov function and the theorem is fulfilled.  $\square$

So, we have:

$$\Delta V_k = V_{k+1} - V_k = \sum_{s=1}^{\mu} \lambda_s [x_{k+1}^T P_s x_{k+1}] - \sum_{s=1}^{\mu} \lambda_s [x_k^T P_s x_k]$$

By reordering and set the matrix  $V_k^s = x_k^T P_s x_k$

$$\begin{aligned} \Delta V_k &= \sum_{s=1}^{\mu} \lambda_s L_s \\ L_s &= V_{k+1}^s - V_k^s \end{aligned} \tag{26}$$

It is enough to show that:

$$L_s < 0 \quad s = 1, \dots, \mu \tag{27}$$

Substituting  $V_{k+1}^s$  and  $V_k^s$  in (26) we have:

$$L_s = \left[ \sum_{i \in S_s} \sum_{j \in S_s} \psi_i \psi_j \Lambda_{ij} x \right]^T P_s \left[ \sum_{i \in S_s} \sum_{j \in S_s} \psi_i \psi_j \Lambda_{ij} x \right] - x^T P_s x \tag{28}$$

$$L_s \leq x^T \left[ \sum_{i \in S_s} \sum_{j \in S_s} \psi_i \psi_j \left[ \Lambda_{ij}^T P_s \Lambda_{ji} - P_s \right] \right] x \tag{29}$$

$$\begin{aligned} &\leq x^T \left[ \sum_{i \in S_s} \psi_i^2 \left[ (\Lambda_{ii} + \Lambda_{ii})^T P_s (\Lambda_{ii} + \Lambda_{ii}) - P_s \right] + \right. \\ &\quad \left. \sum_{i \in S_s} \sum_{j \in S_s}^{j < i} \psi_i \psi_j \left[ \frac{1}{2} (\Lambda_{ij} + \Lambda_{ji})^T P_s (\Lambda_{ij} + \Lambda_{ji}) - 2P_s \right] \right] x \tag{30} \\ &L_s < 0 \rightarrow \Delta V_k < 0 \end{aligned}$$

The first term in (30) is negative definite by (22). The second term is negative definite by (23). Thus, the definite positive quadratic function (24) is a Lyapunov function for the fuzzy control (17), this implicates asymptotically stability in the large. The proof of the theorem is complete.

## 4 Scheduling

In this section, the scheduling theory is used to design an online feedback scheduler for the NCS based on the performance of the control system and the load conditions of the communication network. The scheduler adjusts the sampling/actuation period in function of the QoS and the QoC, the next period is selected like a multiple of the period base ( $h$ ) between the lower and upper bounds obtained by an analysis of the process. The idea is to keep the deadline rate and the system performance at the reference level by adjusting the transmission period. It is proposed a local fuzzy scheduler for each sensor node present in the communication network, which based on external traffic, perform a dynamic scheduling also called feedback scheduling.

The codesign controller/scheduling has as aim to minimize the effects of the time delays and lost packets using the fuzzy controller, while the fuzzy scheduling minimizes the transmission messages without degrading the process performance, modifying the sampling/actuation period into the range that the fuzzy controller guaranteed the stability.

The proposed dynamic scheduling has the configuration shown in Figure 1, its behaviour is as follows: the sensor node sends packets with the system information and its execution period. The controller node adds the error and control signal to the packet and sends it to the actuator node. The actuator computes the system performance and the deadline rate each scheduling period and sends it to the sensor node. Finally, the sensor node modifies the period  $h^+$  based on the information of the deadline rate and the system performance. A deadline occurs when a packet is upgraded after a time limit  $h_{max}$  or when the packet is lost.

In terms of control, the manipulated variable is the actuation period and the controlled variables are the deadline rate and system performance.

The actuator node module calculates the system operation  $\Delta e$  through the mean absolute error (MAE) of  $n$  received packets with a scheduling period  $\delta$ . While the deadline rate  $\Delta h$  is calculated with the  $m$  packets received that have exceeded their deadline  $h_{max}$  plus the lost packets  $\xi$  during the scheduling period.

$$\begin{aligned}\Delta e &= \frac{\sum_{k \in n} |e_k|}{n} \\ m &= \{\forall k \in n | \tau_k > h_{max}\} + \xi \\ \Delta h &= \frac{m}{\delta} \sum_{k \in n} h_k\end{aligned}\tag{31}$$

The deadline rate as a controlled variable is a common metric for the quality of service (QoS). from a real-time point of view is also an important factor that degrades the quality of control (QoC). The QoS controls the number of lost deadlines at an acceptable low level.

In addition, The  $\Delta h$  as a controlled variable can simultaneously address the problems of variable time delays and lost packets. When  $\Delta h$  is kept at a low level, the delays of most packets are less than the deadline and the number of packets lost is limited. As a consequence, the impact of the delay and the lost packets in the QoC is minimized.

The sampling period affects the lost deadline rate, with short sampling periods increasing network utilization, which inherently causes that the network imperfections increase and vice versa. With heavy traffic load, the probability of collisions between nodes is greater, which potentially increases both the time delay and the lost packets and at the same time the lost

deadline rate. On the other hand, a largely lost deadline rate can generally be reduced by increasing the sampling period, particularly when the network is overloaded.

According to the control theory of sampled data, short sampling periods generate better QoC. In this context, QoC can be improved by increasing the utilization efficiency by adjusting the sampling period. This justifies the choice of the sampling period as a manipulated variable that is adjusted with respect to the network conditions, where variations in the unpredictable and dynamic traffic load in the NCS can be effectively compensated.

With the metric of the QoC ( $\Delta e$ ) and QoS ( $\Delta h$ ), the fuzzy scheduler is designed as a control problem, where the metrics form the antecedent part and the consequent part are parallel feedback matrices. The  $i$ -th fuzzy rule has the form of:

$$\text{if } \Delta e \text{ is } \beta_{1j} \text{ and } \Delta h \text{ is } \beta_{2j} \text{ then } \bar{h} = Fz \quad (32)$$

where  $\beta$  are the membership functions and  $z \in \mathbb{R}^2$  with  $z = [\Delta e \ \Delta h]$ ,  $F$  is the feedback scheduling matrix. the overall fuzzy scheduler is:

$$\bar{h} = \sum_{i=1}^M \prod_{j=1}^2 \beta_{ij} F_i z \quad (33)$$

where  $M$  is the number of fuzzy rules. The new sampling period is assigned in the interval:

$$h^+ = \{h_{min} \leq \bar{h} \leq h_{max}\} \quad (34)$$

Given the absence of a mathematical model that describes the relationship between the lost deadline rate, the EAM and the sampling period, the bounds and the fuzzy scheduler are determined based on experimentation.

#### 4.1 Analytical codesign

If the fuzzy model is locally controllable, i.e.  $(\Phi_i, \Gamma_i)$ ,  $i = 1, \dots, l$ , are controllable pairs, the feedback control matrices  $K_i$ , ( $i = 1, 2, \dots, l$ ) can be calculated using eq. (18), with matrices  $R_i$  and  $Q_i$  to get a desired performance.

The procedure of the fuzzy codesign is as follows:

Step 1: Define the intervals of the compound time based on the measurements of the network to define the bounds of the compound time.

Step 2: For the plant (eq. 1), define the  $h_i$  actuation periods for each  $F_i$  matrix in the fuzzy scheduler. Define the membership functions for the control error  $\Delta e$  and deadline rate  $\Delta h$ .

Step 3: Discretize the model (eq. 1) for each  $h_i$  to set  $\Gamma_i$  and  $\Phi_i$  in the fuzzy controller.

Step 3: Verify that all local discrete systems are controllable. That is,  $rank(\Gamma_i, \Phi_i \Gamma_i, \dots, \Phi_i^{n-1} \Gamma_i) = n$ ,  $i = 1, \dots, l$ .

Step 4: Calculate the feedback matrices  $K_i$  of each local system via LQR assigning the matrices  $R_i$  and  $Q_i$  according to the desired performance (Eq. 18).

Step 5: Find all the  $\mu$  regions with almost two overlap rules and apply Theorem 4.1 to check the stability of the closed-loop. If the system is not stable, go back to step 4 to reassign the matrices  $R_i$  and  $Q_i$ .

## 5 Experiments and results

The fuzzy codesign approach is tested in a nonlinear MIMO system with an Ethernet network, the experiment generated variable traffic into the network with external nodes. The performance is compared with a Hybrid controller [11].

## 5.1 Case study

The case study is a MIMO nonlinear, open-loop unstable and time-varying system. It is a 2-DOF helicopter system integrated to an Ethernet network (Figure 5), detailed information can be found in [19]. The sensor and actuator nodes are Pentium 4 with 1028 Mb RAM with an INTEL 10/100 Mb Ethernet card, each has an XPC target operative systems and are connected through a switch, the controller is an embedded system with a microcontroller board based on the ATmega32u4 and the Atheros AR9331. The Atheros processor uses Linino OS a Linux distribution. The board has built-in Ethernet slot, a 16 MHz crystal oscillator. The sensor node has an A/D 10 bits resolution and the actuator has a D/A 8 bits resolution. The tasks for the NCS are set as follows,  $5\text{ ms}$  as minimum sampling period of sensor and actuator nodes, the controller node is driven by event.

The experiment consists of a 2D helicopter simulator mounted on a fixed base with two propellers that are driven by DC motors as is shown in Figure 5. The front propeller controls the elevation of the helicopter nose about the pitch axis and the back propeller controls the motion about the yaw axis. The pitch and yaw angles are measured using high-resolution encoders.

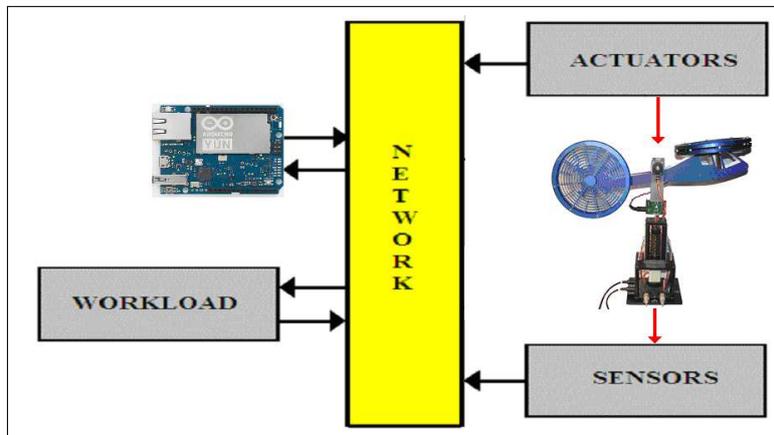


Figure 5: Networked Control System with workload

The Euler-Lagrange method is used to derive the nonlinear equations describing the motions of the helicopter [19]. From its nonlinear equations of motions, the linear continuous state space models with  $x = [\theta \ \psi \ \dot{\theta} \ \dot{\psi}]$  and  $u = [V_p \ V_y]$  are

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned} \tag{35}$$

$$\begin{aligned}
A &= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\frac{B_p}{J_p+ml^2} & 0 \\ 0 & 0 & 0 & -\frac{B_y}{J_y+ml^2} \end{bmatrix} \\
B &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{K_{pp}}{J_p+ml^2} & \frac{K_{py}}{J_p+ml^2} \\ \frac{K_{yp}}{J_y+ml^2} & \frac{K_{yy}}{J_y+ml^2} \end{bmatrix} \\
C &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}
\end{aligned} \tag{36}$$

where  $K_{pp}$  is the thrust torque acting on pitch axis from pitch motor,  $K_{py}$  the thrust torque acting on yaw axis from pitch motor,  $K_{yp}$  the thrust torque acting on pitch axis from yaw motor,  $K_{yy}$  the thrust torque acting on yaw axis from yaw motor,  $B_p$  the viscous damping coefficient about pitch axis,  $B_y$  the viscous damping coefficient about yaw axis,  $l$  the center of mass length along helicopter bod,  $g$  the gravity constant,  $m$  the total moving mass of the helicopter,  $J_p$  the total moment of inertia about pitch pivot,  $J_y$  the total moment of inertia about yaw pivot,  $V_p$  the voltage of pitch motor,  $V_y$  voltage of yaw motor,  $\theta$  the angle about pitch axis and  $\psi$  the angle about yaw axis.

## 5.2 Results

The fuzzy model is obtained discretizing the helicopter model (Eqs. (35)-(36)) with different sampling-actuation periods  $h_j = [5, 10, 15, 20] \text{ ms}$ , the compound time is previously analyzed generating the bounds  $v = [5 - 300] \text{ ms}$ . The fuzzy model has four rules with the antecedent parameters

$$\begin{aligned}
\rho &= [5, 10, 15, 20] \times 10^{-3} \\
\sigma &= [12, 24, 30, 30] \times 10^{-4}
\end{aligned}$$

The fuzzy controller is designed by LQR method and the Lyapunov conditions are fulfilled (22)-(23).

$$K = \begin{bmatrix} 15.4 & 1.53 & 4.91 & 0.677 & 12.2 & 0.718 \\ -1.97 & 17.3 & -0.241 & 6.2 & -1.24 & 7.03 \end{bmatrix}$$

The fuzzy scheduler is designed with the sampling-actuation periods  $h_j = [5, 10, 15, 20] \text{ ms}$ . The range of the lost deadlines is  $\Delta e = [0, 1]$ , the MAE is  $\Delta h = [0, 1.4]$  and the scheduling period  $\delta = 1 \text{ s}$ .

The network imperfections and the system performance were measured to show the codesign approach with a square reference. The overall system performance is compared with a hybrid control designed to be stable in the compound time range.

The network behaviour with medium traffic generates a peculiar behaviour in the delay and the loss of deadlines (6). The time delay increases at the start of traffic with a maximum of 31 ms and average of 5 ms, but as the traffic increases, delays become deadlines with a 5.8% and average of 132 ms between missed deadlines. The upper graph shows the time delay where the average traffic starts in 50 s, however, the effect starts seconds later and decreases around the second 55, giving rise to an increase in the lost deadlines (80 s), but its effect extends to the 85s because the switch continues to empty the queue.

In this case, the fuzzy scheduler change the assignment policy for the sampling-actuation period (7) (inf.), When medium traffic appears, the EAM (sup.) had variations greater than 0.1,

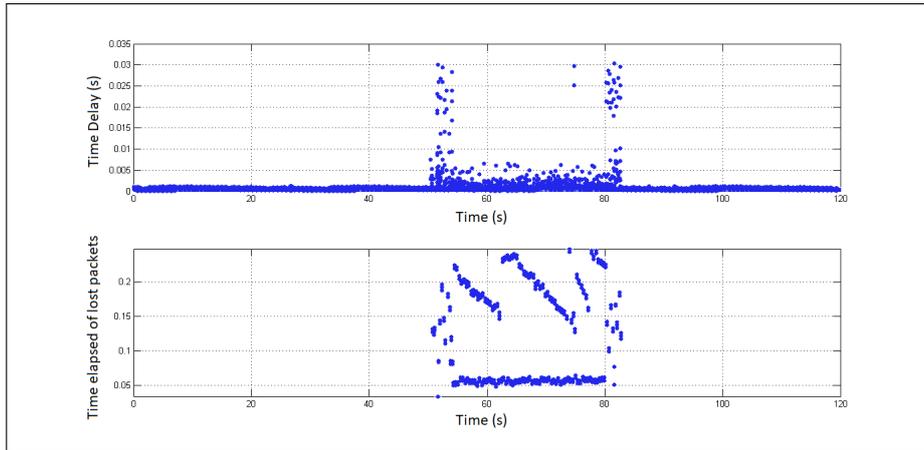


Figure 6: Time delays and time elapsed between lost packets with medium traffic (50 - 80 s).

which represents an error in steady state greater than 2% , and added to the presence of 5.8 of lost deadlines (med.), the fuzzy scheduler decides to decrease the sampling period slowly but without overloading the network and gradually improve the performance of the system.

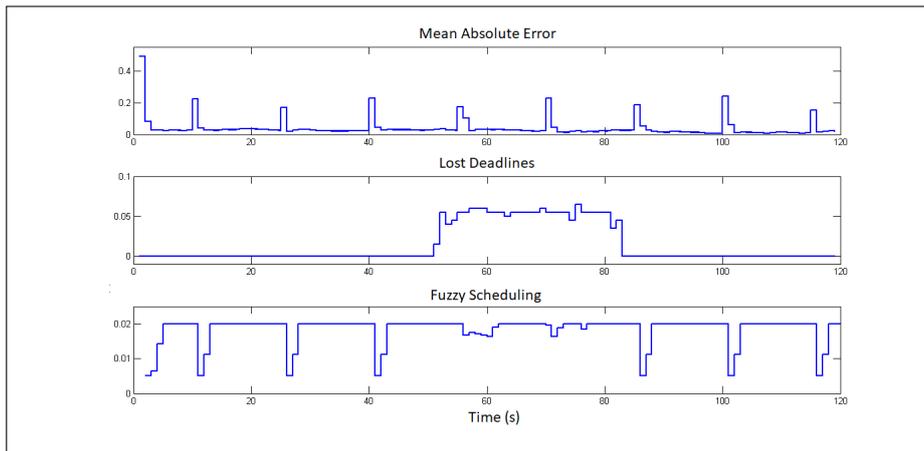


Figure 7: MAE, lost deadlines and the sampling-actuation period with medium traffic.

With medium traffic, the fuzzy codesign and the hybrid control remain stable (Fig. 8), but the fuzzy codesign (sup.) is the one that fulfils the control criteria. The codesign (sup.) with medium traffic has an overshoot of  $\zeta = 15^\circ$ , with a stable steady error  $e_s s = 0.5$  and an setting time  $T_s = 4$  s.; while the hybrid control (inf.) degrades its performance against average traffic with an overshoot of  $\zeta = 34^\circ$ , with a stable steady error  $e_s s = 6$  and with oscillations. At this level of traffic, the hybrid control ensures system stability but with poor performance.

In the case of heavy traffic (Figure 3), the time delay and the lost deadlines were increased, the maximum delay was 30 ms, with an average of 12 ms, and deadlines were 7.3% and the average time of 122 ms between lost deadlines. The heavy traffic started at 50 s, however, its effect was visible until the 85 s even when the traffic was finished in the 80 s.

Figure 9 shows the pitch position for both the fuzzy codesign and the hybrid controller. In case of the heavy traffic measured (Figure 3), the hybrid controller has an erratic behavior with an overshoot  $\zeta = 36^\circ$ , a steady stable error  $e_s s = 10$  and setting time  $T_s > 9$  s. While, the fuzzy codesign curves indicate a stable behavior with an overshoot  $\zeta = 30^\circ$ , a steady stable error

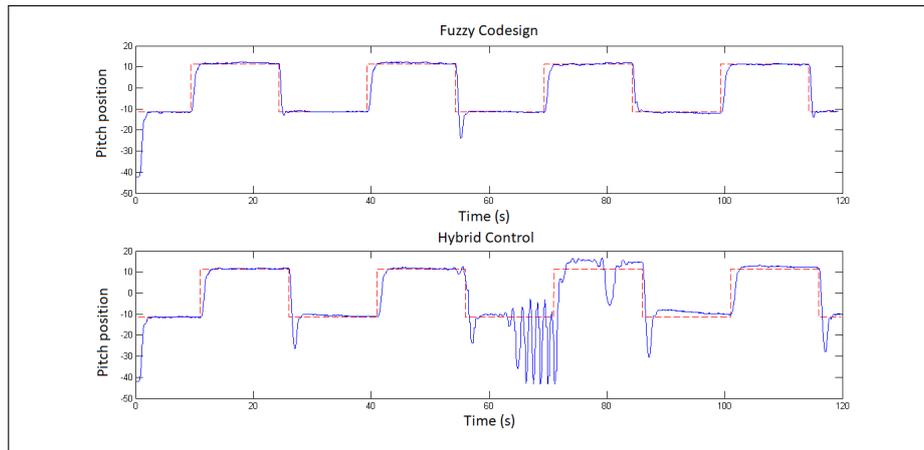


Figure 8: Comparison between the codesign and an hybrid controller with medium traffic.

$e_s s = 0.8$  and setting time  $T_s = 2$  s.

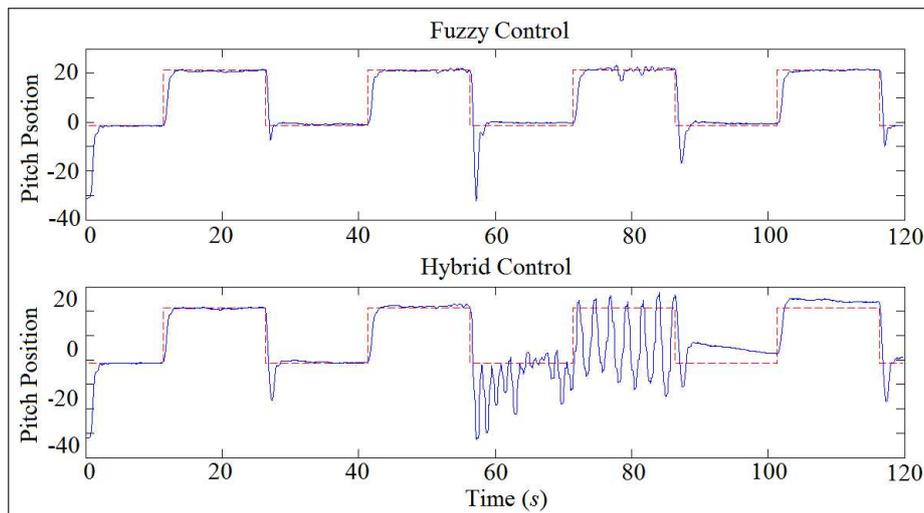


Figure 9: Comparison between the codesign and an hybrid controller with heavy traffic.

The fuzzy scheduler changes the sampling period (Figure 10), Because the MAE (sup.) has variations greater than 0.25 representing a steady-state error greater than 5%, And added to the presence of 7.8% of lost deadlines (med.), The fuzzy scheduler decides to slow down the sampling period in order to correct the error even when traffic on the network increases. This demonstrates that the fuzzy codesign can apply a dynamic control dependent on the network behaviour with a stable design in all the range of  $v$ .

## 6 Conclusion

A fuzzy codesign approach was presented to minimize the effects of the network-induced imperfections. The approach was designed with a controller and scheduler together in function of the network imperfection measurements and the continuous model of the system. The fuzzy controller is designed to select the control signal depending on the compound time estimation, with multiple discrete model that represents the process in function of the actuation periods dis-

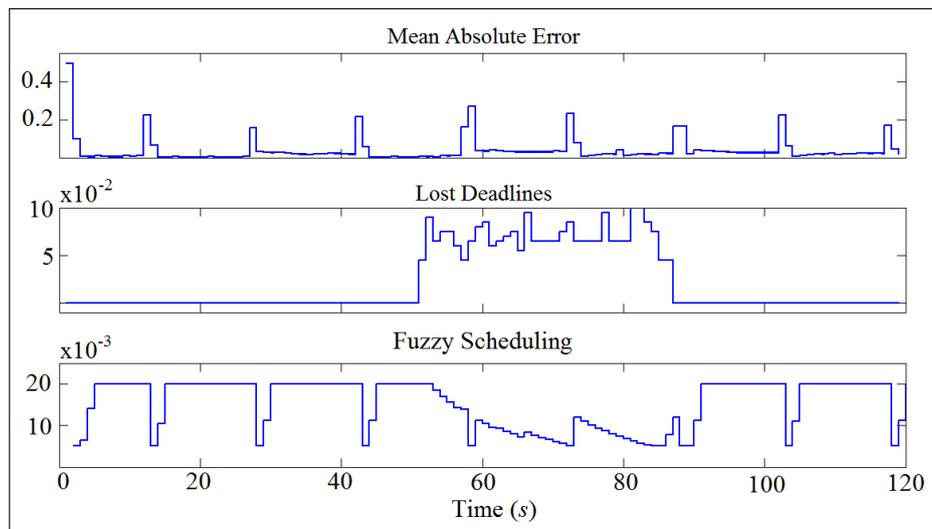


Figure 10: MAE, lost deadlines and the sampling-actuation period with heavy traffic.

appeared the sampling and communication jitter. The fuzzy scheduler was designed to control the sampling-actuation period in function of the system performance and the behaviour of the network. The codesign approach is applied to a nonlinear, time-varying MIMO system interconnected with an Ethernet network and employing an embedded system as the controller-scheduler. The codesign performance was compared with a hybrid control designed to counteract the effects of delay in the same range as codesign. The fuzzy codesign had the best performance within the entire range of compound time.

### Acknowledgement

The authors would like to thank DCC-IIMAS-UNAM for the support and for their technical support to Sandra Sauza Escoto and Joaquin Morales Rosales. This work was supported by UNAM-PAPIIT IA104218.

### Bibliography

- [1] Albertos, P.; Crespo, A. (1999); Real-time control of non-uniformly sampled systems, *Control Engineering Practice*, 7(4), 445-458, 1999.
- [2] Aström, K.J. (2007); *Event Based Control, In Analysis and Design of Nonlinear Control Systems: In Honor of Alberto Isidori*, Springer Verlag, 2007.
- [3] Benitez-Perez, H. et al. (2016); A Fuzzy Networked Control System Following Frequency Transmission Strategy, *International Journal of Computers Communications & Control*, 11(1), 11-25, 2016.
- [4] Chae, S.; Nguang, S. K. (2014); SOS based robust  $H_\infty$  fuzzy dynamic output feedback control of nonlinear networked control systems, *IEEE Trans. Cybern.*, 44(7), 1204-1213, 2014.
- [5] Chai, S.; Liu, G.P.; Rees, D.; Xia Y. (2008); Design and practical implementation of internet-based predictive control of a servo system, *IEEE Trans. on Control Systems Technology*, 16(1), 158-168, 2008.

- 
- [6] Chow, M.-Y.; Tipsuwan, Y. (2003); Gain adaptation of networked DC motor controllers based on QoS variations, *IEEE Trans. on Industrial Electronics*, 50(5), 936-943, 2003.
- [7] Du, D. (2012); Reliable  $H_\infty$  control for Takagi-Sugeno fuzzy systems with intermittent measurements, *Nonlinear Analysis: Hybrid Systems*, 6(4), 930-941, 2012.
- [8] Guan, X.P.; Chen, C. L. (2004); Delay-dependent guaranteed cost control for TS fuzzy systems with time delays, *IEEE transactions on fuzzy systems*, 12(2), 236-249, 2004.
- [9] Hespanha, J.; Naghshtabrizi, P.; Xu, Y. (2007); A survey of recent results in networked control systems, *Proc. IEEE*, 138-162, 2007.
- [10] Hu, H.; Liu, G.; Rees, D. (2007); Event-driven networked predictive control, *IEEE Transactions on Industrial Electronics*, 54(3), 1603- 1613, 2007.
- [11] Ji, K.; Kim, W.J. (2007); Stochastic optimal control and network co-design for networked control systems, *International Journal of Control, Automation and Systems*, 5(5), 515-525, 2007.
- [12] Ji, K.; Kim, W.J. (2007); Robust control for networked control systems with admissible parameter uncertainties, *International Journal of Control, Automation and Systems*, 5(4), 372-378, 2007.
- [13] Li, H.; Sun, Z.; Liu, H.; Chow, M.Y. (2008); Predictive observer-based control for networked control systems with network-induced delay and packet dropout, *Asian Journal of Control*, 10(6), 1-3, 2008.
- [14] Li, K.; Baillieul, J. (2004); Robust quantization for digital finite communication bandwidth (DFCB) control, *IEEE Trans. on Automatic Control*, 49(9), 1573-1584, 2004.
- [15] Li, H.; Wu, C.; Feng, Z. (2015); Fuzzy dynamic output-feedback control of non-linear networked discrete-time system with missing measurements, *IET Control Theory and Applications*, 9(3), 327-335, 2015.
- [16] Lozoya, C.; Marti, P.; Velasco, M.; Fuertes, J.M. (2008); Control Performance Evaluation of Selected Methods of Feedback Scheduling of Real-time Control Tasks, *17th IFAC World Congress*, July, 2008.
- [17] Marti, P.; Velasco, M. (2007); Toward Flexible Scheduling of Real-Time Control Tasks: Reviewing Basic Control Models, *10th International Conference on Hybrid Systems, Computation and Control, LNCS*, 2007.
- [18] Mendez-Monroy, P.E.; Velasco, M.; Fuertes, J.M.; Benitez-Perez, H. (2012); Fuzzy observer based Fault Detection for Network Control Systems with Periodic Actuation Tasks, *8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes (SafeProcess 2012)*, 2012.
- [19] Mendez-Monroy, P. E.; Benitez-Perez, H. (2011); Fuzzy control with estimated variable sampling period for non-linear networked control systems: 2-DOF helicopter as case study, *Transactions of the Institute of Measurement and Control*, 34(7), 802-814, 2011.
- [20] Mendez-Monroy, P. E. (2017); Fuzzy codesign for networked control systems, *14th International Conference on Electrical Engineering, Computing Science and Automatic Control*, DOI: 10.1109/ICEEE.2017.8108887, 2017.

- 
- [21] Montestruque, L.A.; Antsaklis, P.K. (2007); Static and dynamic quantization in model-based networked control systems, *International Journal of Control*, 80(1), 87-101, 2007.
- [22] Nilsson, J.; Bernhardsson, B.; Wittenmark, B. (1998); Stochastic analysis and control of real-time systems with random time delays, *Automatica*, 34(5), 57-64, 1998.
- [23] Park, J.; Kim, J.; Park, D. (2001); LMI-based design of stabilizing fuzzy controllers for nonlinear systems described by Takagi-Sugeno fuzzy model, *Fuzzy Sets and Systems*, 122, 73-82, 2001.
- [24] Peng, C.; Yang, T.C. (2010); Communication-delay-distribution-dependent networked control for a class of T-S fuzzy systems, *IEEE Transaction on Fuzzy Systems*, 18(2), 326-335, 2010.
- [25] Seiler, P.; Sengupta, R. (2005); An  $H_\infty$  approach to networked control, *IEEE Transaction on Automatic Control*, 50(3), 356-364, 2005.
- [26] Tanaka, K.; Wang, H. O. (2001); *Fuzzy Control Systems Design and Analysis: A Linear Matrix Inequality Approach*, Wiley & Sons, Inc., 2001.
- [27] Tipsuwan, Y.; Chow, M.Y. (2003); Control methodologies in networked control systems, *Control Engineering Practice*, 11, 1099-1111, 2003.
- [28] Tipsuwan, Y.; Chow, M. Y. (2004); On the Gain Scheduling for Networked PI Controller Over IP Network, *IEEE/ASME Transaction Mechatronics*, 9(3), 491-498, 2004.
- [29] Tong, S.W.; Qian, D.W.; Liu, G.P. (2014); Networked Predictive Fuzzy Control of Systems with Forward Channel Delays Based on a Linear Model Predictor, *International Journal of Computers Communications & Control*, 9(4), 471-481, 2014.
- [30] Yi, H.C.; An C.J.; Choi J.Y. (2017); Compensation of Time-Varying Delay in Networked Control System over Wi-Fi Network, *International Journal of Computers Communications & Control*, 12(3), 415-428, 2017.
- [31] Yu, M.; Wang, L.; Chu, T.G., Xie, G.M. (2004); Stabilization of networked control systems with data packet dropout and network delays via switching system approach, *Proc. of the 43th IEEE Conference on Decision and Control*, 3539-3544, 2004.
- [32] Zhang, W.; Branicky, M.S.; Phillips, S.M. (2001); Stability of networked control systems, *IEEE Cont. Sys. Mag.*, 21(1), 84-99, 2001.
- [33] Zhang, H.; Duan, G.; Xie, L. (2007); Linear quadratic regulation for linear time-varying systems with multiple input delays, *Automatica*, 42(9), 1465-1476, 2007.
- [34] Zhao, Y.; Gao, H.; Chen, T. (2010); Fuzzy constrained predictive control of non-linear systems with packet dropouts, *IET control theory and applications*, 4(9), 1665-1677, 2010.

# EPAK: A Computational Intelligence Model for 2-level Prediction of Stock Indices

L. Tang, H. Pan, Y. Yao

## Li Tang

1. School of Management and Economics  
University of Electronic Science and Technology of China  
Chengdu 611731, China  
2. Tianfu College of Southwestern University of Finance and Economics  
Chengdu 610052, China  
tangli@std.uestc.edu.cn

## Heping Pan\*

Intelligent Finance Research Center  
Chongqing Institute of Finance  
Chongqing 400067, China  
\*Corresponding author: panhp@swingtum.com

## Yiyong YAO

Tianfu College of Southwestern University of Finance and Economics  
Chengdu 610052, China  
yiyongyao@yahoo.com

**Abstract:** This paper proposes a new computational intelligence model for predicting univariate time series, called EPAK, and a complex prediction model for stock market index synthesizing all the sector index predictions using EPAK as a kernel. The EPAK model uses a complex nonlinear feature extraction procedure integrating a forward rolling Empirical Mode Decomposition (EMD) for financial time series signal analysis and Principal Component Analysis (PCA) for dimension reduction to generate information-rich features as input to a new two-layer K-Nearest Neighbor (KNN) with Affinity Propagation (AP) clustering for prediction via regression. The EPAK model is then used as a kernel for predicting each of all the sector indices of the stock market. The sector indices predictions are then synthesized via weighted average to generate the prediction of the stock market index, yielding a complex prediction model for the stock market index. The EPAK model and the complex prediction model for stock index are tested on real historical financial time series in Chinese stock index including CSI 300 and ten sector indices, with results confirming the effectiveness of the proposed models.

**Keywords:** empirical mode decomposition, principal component analysis, affinity propagation, k-nearest neighbor, time series, stock index prediction.

## 1 Introduction

Time series prediction is a practical issue. Especially the financial time series prediction with important economic significance has attracted serious attention from both finance and computer science researchers. A fair large of literatures research on financial prediction methods[23], typically including ARIMA [8,27], ARCH [3,13],GARCH [1,16], chaos-theoretical [17,18], Artificial Neural Network (ANN) [10,20], Support Vector Machine (SVM) [9,19], and K-Nearest Neighbor (KNN) [12,28].

From these equivalent researches, it is realized that the foremost key factor for effective prediction is the feature extraction which should generate essential information effectively. Practically, the financial time series feature extraction is equivalent to the signal analysis. Thus we

can apply a signal processing technique such as the Empirical Mode Decomposition [6] which is an effective method used widely in dealing with nonlinear and non-stationary signals [21,25]. However, the end effect of EMD [15] has not been considered usually in most of the researches except the forward rolling EMD with a sliding window proposed by Zhang and Pan in 2015[26]. Through a forward rolling EMD, the original time series is decomposed into multilevel IMFs with a high dimension, thus the Principal Components Analysis (PCA) [11] can be applied for dimension reduction. Generally, we propose a nonlinear feature extraction procedure integrating a forward rolling EMD and PCA in this paper.

For prediction modeling, KNN is a nonparametric algorithm which can predict via nonlinear regression. However, it should be note that KNN has a large amount of calculations and a large prediction deviation may be occurred when the samples are in disequilibrium. To tackle these issues, this paper proposes a two-layer KNN with Affinity Propagation (AP) clustering. AP is an effective clustering method with sensory signals process and data patterns detection [5]. Integrate the nonlinear feature extraction procedure and a two-lay KNN, this paper proposes a new computational intelligence model for univariate time series prediction called EPAK and construct a complex prediction model for stock market index synthesizing all the sector index predictions applying EPAK as a kernel.

## 2 An EMD-PCA-AP-KNN (EPAK) financial prediction model

### 2.1 Frame work of an EPAK financial prediction model

The EPAK model which is a computational intelligence financial prediction model needs to learn, adapt to and evolve along with changing financial situations. In General, we should first define a time frame and assume a historical financial time series long enough exists. In this paper, we focus on a daily time frame thus we can express a financial price time series on day  $t$  as

$$X(t) = (O(t), H(t), L(t), C(t), V(t)), \tag{1}$$

where  $O(t), H(t), L(t), C(t)$  and  $V(t)$  are the open price, high price, low price, close price and volume. In this paper we only consider the close price, so  $C(t)$  is set as  $X(t)$ . We can define a relative return of the price for  $X(t)$  as

$$R(t, \lambda) = \frac{X(t) - X(t - \lambda)}{X(t - \lambda)}, \tag{2}$$

where  $\lambda$  is the prediction step length of time series with a basic step length of  $\lambda = 1$ . Without any other specification, we use  $R(t)$  as  $R(t, \lambda)$ . Therefore, we can express a historical relative return data set as

$$DR(t, T) = (R(t - (T - w) + 1), \dots, R(t + 1), R(t)), \tag{3}$$

where  $T$  is the number of total days and  $w \ll T$  is the width of a sliding window applied for intercepting the historical data set.

In general, an EPAK financial prediction model can be expressed as

$$EPAK : P(t) \implies AKNN(k) \implies R(t + \lambda), \tag{4}$$

or mathematically

$$EPAK : R(t + \lambda) = AKNN(P(t), k), \tag{5}$$

where AKNN stands for a two-layer KNN with AP,  $P(t)$  means the principal components generated by PCA,  $k$  is the structural parameter, and  $R(t + \lambda)$  is the prediction output. For details, the EPAK model works through two processes, first the feature extraction integrating EMD for signal decomposition of financial time series and PCA for dimension reduction, second the prediction procedure applying a two-layer KNN regression with AP. Thus we can re-express the EPAK model as

$$DR(t, T) \Rightarrow EMD \Rightarrow PCA \Rightarrow P(t) \Rightarrow AP \Rightarrow AKNN(k) \Rightarrow R(t + \lambda), \quad (6)$$

or mathematically

$$R(t, \lambda) = AKNN\{AP[FE(PCA(EMD(DR(t, T))))]\}, \quad (7)$$

where  $FE()$  denotes the feature extraction process.

## 2.2 A nonlinear feature extraction process integrating a forward rolling EMD and PCA

As a foremost key process for financial time series prediction, feature extraction should concentrate essential information from the historical data set and input to the prediction model. The EPAK model encompasses a complex nonlinear feature extraction procedure special for financial time series, integrating a forward rolling EMD and PCA.

### A forward rolling EMD for financial time series

To start a feature extraction process, we apply the forward rolling EMD for signal decomposing on the historical data set  $DR(i - 1, T), i = t, \dots, t - (T - w) + 1$  and make it as the input,  $R(t), i = t, \dots, t - (T - w) + 1$  as the output of prediction. Therefore, the input-output data set can be expressed as

$$DP(t, T - w) = \begin{pmatrix} EMD(DR(t - 1), T) & \rightarrow & R(t) \\ EMD(DR(t - 2), T) & \rightarrow & R(t - 1) \\ \dots & \dots & \dots \\ EMD(DR(t - (T - w)), T) & \rightarrow & R(t - (T - w) + 1) \end{pmatrix}. \quad (8)$$

After EMD, the original data set has been decomposed to multilevel IMFs which can satisfy two conditions according to IMF definition [6]: 1) for a whole series, the total number of extrema and zero-crossing points should be equal or differ no more than one, 2) for any time period, the mean of the upper envelope and the lower envelope that are formed by local maxima and minima equals zero.

Take the EMD decomposition of  $DR(t)$  as an example, the EMD decomposition consists of three parts:

1) Sifting process:

- Calculate all the local maxima and minima of  $DR(t)$ .
- Generate upper and lower envelopes with maxima and minima by cubic spline, and calculate the mean value

$$m_i(t) = \frac{(ue(t) + le(t))}{2}, \quad (9)$$

where  $i = 1, 2, \dots$  indicates the  $i$ th-order,  $ue(t)$  and  $le(t)$  are the upper and lower envelopes.

- Generate a detail component of decomposition

$$h_i(t) = dr(t) - m_i(t), \tag{10}$$

when  $i = 1, dr(t) = DR(t)$ .

- 2) IMF checking: check whether  $h_i(t)$  can satisfy the two conditions of IMF definition:

- If it can,  $h_i(t)$  is an IMF, and the residual is

$$r(t) = dr(t) - c(t), \tag{11}$$

$$c(t) = h_i(t). \tag{12}$$

Let  $dr(t) = r(t)$  and continue to next sifting.

- If it cannot, let  $dr(t) = h_i(t)$  and continue to next sifting by 1).

- 3) Sifting stop:

- To ensure the instantaneous frequency defined by IMF has sufficient physical significances, a metric SD is defined by Huang [6] to determine whether to stop sifting. SD can be expressed as

$$SD = \sum_{i=1}^N \left[ \frac{|h_{i+1}(t) - h_i(t)|^2}{(h_i(t))^2} \right], \tag{13}$$

when  $0.2 < SD < 0.3$ , stop sifting.

- If the number of extrema including maxima and minima of  $r(t)$  is smaller than two, stop sifting.

- Return all the results as

$$DR(t) = \left( \sum_{i=1}^n c_i \right) + r, \tag{14}$$

or in more detail

$$DR(t) = \begin{pmatrix} IMF_1(t, w) \\ \vdots \\ IMF_n(t, w) \end{pmatrix} + r. \tag{15}$$

Equation (15) means that EMD can decompose the original data set into  $n$  (generally, let  $n \leq 5$ ) multilevel IMFs and a residual  $r$ , thus Eq.(8) can be rewritten as

$$DP(t, T - w) = \{ \mathbf{D} \rightarrow \mathbf{R} \}, \tag{16}$$

$$\mathbf{D} = \begin{pmatrix} IMF_1(t - 1, w) & \cdots & IMF_n(t - 1, w) \\ IMF_1(t - 2, w) & \cdots & IMF_n(t - 2, w) \\ \vdots & \vdots & \vdots \\ IMF_1(t - (T - w), w) & \cdots & IMF_n(t - (T - w), w) \end{pmatrix}, \tag{17}$$

$$\mathbf{R} = \begin{pmatrix} R(t) \\ R(t - 1) \\ \vdots \\ R(t - (T - w) + 1) \end{pmatrix}. \tag{18}$$

It should be note that each row of  $\mathbf{D}$  is high dimensional since it consists of multilevel IMFs time series. Therefore, a PCA algorithm can be applied for reducing the high dimension.

### PCA for dimension reduction

Compared to other methods of dimension reduction such as Linear Discriminant Analysis (LDA), Locally Linear Embedding (LLE), and Laplacian Eigenmaps (LE), PCA can maintain the information contained in the original data as much as possible after dimension reduction. However, for financial time series prediction, the feature extraction needs to keep the more information of original data the better robustness and prediction accuracy of the model. Therefore, PCA is used for dimension reduction and feature extraction in this paper. PCA reduces data dimension by an orthogonal linear transformation which is actually a singular value decomposition process. Thus we can express a PCA algorithm as three processes:

1) Form a matrix consists of principal components: normalize the high-dimensional matrix  $\mathbf{D}$  to  $\mathbf{Y}$

$$\mathbf{Y} = (\text{normalization}(\mathbf{D}))^T. \quad (19)$$

Apply a singular value decomposition of  $\mathbf{Y}$

$$\mathbf{Y} = \mathbf{V}\mathbf{\Sigma}\mathbf{W}^T, \quad (20)$$

where  $\mathbf{V}$  and  $\mathbf{W}$  are orthogonal matrices formed by the eigenvectors of  $\mathbf{Y}\mathbf{Y}^T$  and  $\mathbf{Y}^T\mathbf{Y}$  respectively,  $\mathbf{\Sigma}$  is a nonnegative rectangular diagonal matrix whose left part consists of the eigenvalues  $\lambda(i), i = 1, 2, \dots, p$  of  $\mathbf{Y}\mathbf{Y}^T$ . Thus we can generate a transformed matrix  $\mathbf{P}$  that consists of principal components in turn,

$$\mathbf{P} = \mathbf{S}^T\mathbf{V} = \mathbf{W}\mathbf{\Sigma}^T\mathbf{V}^T\mathbf{V} = \mathbf{W}\mathbf{\Sigma}^T. \quad (21)$$

2) Find a new lower dimension: PCA is used for reducing dimension, thus the dimension of matrix  $\mathbf{P}$  should be reduced. Actually, the first  $r$  in  $l$  principal components concentrate the most essential information of matrix  $\mathbf{P}$ . To find the value of  $r$ , we can use the Cumulative Contribution Rate (CCR) which is generally required to be more than a threshold (such as 85%) [15] for help,

$$CCR_r = \frac{\sum_{i=1}^r \lambda(i)}{\sum_{i=1}^l \lambda(i)} > 85\%, \quad (22)$$

where  $r$  is the new lower dimension and  $r \ll l$ .

3) Generate an information-rich matrix: construct a new matrix  $\mathbf{\Sigma}_r$  as a  $r \times l$  matrix by

$$\mathbf{\Sigma}_r = \mathbf{I}_{r \times l}\mathbf{\Sigma}. \quad (23)$$

Then we can form a lower dimensional matrix  $\mathbf{P}_r$  of  $\mathbf{P}$

$$\mathbf{P}_r = \mathbf{W}(\mathbf{\Sigma}_r)^T, \quad (24)$$

where  $\mathbf{P}_r$  is the new low-dimensional matrix which should be input to a two-layer KNN with AP for prediction.

### 2.3 A two-layer KNN algorithm with AP

KNN is a nonparametric algorithm that can be used for nonlinear regression. To improve the KNN algorithm, this paper proposes a two-layer KNN with AP. We apply AP to transform the feature into clusters as input to a two-layer KNN for prediction.

**AP for clusters generation**

Assume the feature extracted by a forward rolling EMD and PCA comprises  $N$  data points. AP initially regards every data point as a potential cluster center. Then measure the similarity between any data pairs and accumulate evidences for an iterative procedure which can find the final suitable exemplars. For details, we can describe an AP algorithm as three procedures:

1) Construct a similarity matrix: in this paper, we measure the similarity between each data pairs with Euclidean which is one of the classical similarity metric for time series. Thus the similarity can be defined as

$$S_{im}(i, j) = -\|i - j\|^2, \tag{25}$$

where  $i$  and  $j$  are any data pairs of the  $N$  data points. When  $j = i$ ,  $S_{im}(i, i) = p(i)$ , means the preference of point  $i$  can be an suitable exemplar. Therefore,  $p$  is set as a preference parameter which can affect the clustering solutions. The similarity matrix  $S_{N \times N}$  can be composed of all  $S_{im}$  values in order.

2) Find suitable exemplars: an iterative procedure based on  $S_{N \times N}$  can be helpful to find suitable exemplars. In this procedure, the evidences of "Responsibility"  $R$  and "Availability"  $A$  should be accumulated. Assume point  $c$  is the exemplar candidate and  $i$  is any point.  $R(i, c)$  indicates the suitability of  $c$  for being the exemplar of  $i$ , and  $A(i, c)$  indicates the appropriateness of  $i$  for choosing  $c$  as the exemplar,

$$R(i, c) = S_{im}(i, c) - \max_{j \neq c} \{A(i, j) + S_{im}(i, j)\}, \tag{26}$$

$$A(i, c) = \min\{0, R(c, c) + \sum_{j \neq i, c} \max[0, R(j, c)]\}, \tag{27}$$

When  $R(i, c) + A(i, c)$  achieves the maximum,  $c$  is the most suitable exemplar of  $i$ .

3) Choose an optimal clustering solution: after finding the final exemplars, there are several clustering solutions and the optimal one should be chosen. An effective evaluation of clustering solution is the Silhouette Coefficient which can reflect the separability and compactness between clusters [22]. Assume there are clusters  $C_i, i = 1, 2, \dots, n$ ,  $SC$  for any point in cluster  $C_i$  can be expressed as

$$SC(x_i) = \frac{\min[a(x_i, C_j)] - b(x_i)}{\max\{b(x_i), \min[a(x_i, C_j)]\}}, \tag{28}$$

where  $a(x_i, C_j)$  is the mean dissimilarity between  $x_i$  and all the other points in cluster  $C_j, j \neq i, b(x_i)$  is the mean dissimilarity between  $x_i$  and all the other points in cluster  $C_i$ . Calculate the mean value of overall  $SC$  as

$$SC_m = \text{mean}\left\{\sum_{i=1}^N SC(x_i)\right\}, \tag{29}$$

The higher the  $SC_m$  value is, the better the clustering solution [4]. Therefore, we can choose the optimal clustering solution with a highest  $SC_m$ .

**A two-layer KNN**

We propose a two-layer KNN with AP consists of three functions AKNN, AKNN1, and AKNN2. Assume  $x(t + \lambda) = DR(t + \lambda, T)$  is the future prediction point and  $x(t) = DR(t, T)$  as input. AKNN is equivalent to the input-output of a two-layer KNN. For the first layer which can reduce the computation, AKNN calls AKNN1 to find  $x(t)$ 's nearest exemplar  $c_{near}$  and ascribes  $x(t)$  to the same cluster  $C_{near}$ . And for the second layer which can ensure equilibrium samples,

AKNN calls AKNN2 to find the  $k$  nearest neighbors in  $C_{near}$  for prediction. For details, a two-layer KNN can be defined as

**Function AKNN** inputs  $x(t)$  and  $C_i, i = 1, 2, \dots, n$  which are the clusters generated by AP, outputs the prediction  $x(t + \lambda) = DR(t + \lambda, T)$ ,

$$(x(t + \lambda)) = AKNN(x(t), C_i, k). \quad (30)$$

AKNN generates predictions by calling AKNN1 and AKNN2 in turn.

**Function AKNN1** inputs  $x(t)$  and  $C_i, i = 1, 2, \dots, n$ , outputs  $C_{near}$  and its exemplar  $c_{near}$

$$(c_{near}, C_{near}) = AKNN1(x(t), C_i, k = 1), \quad (31)$$

where  $k = 1$  means  $c_{near}$  is  $x(t)$ 's nearest exemplar. Calculate the similarity between  $x(t)$  and each exemplar  $c_i, i = 1, 2, \dots, n$  as

$$S_{im}(x(t), c_i) = -\|x(t) - c_i\|^2. \quad (32)$$

When  $S_{im}(x(t), c_{near})$  achieves the maximum,  $c_{near}$  is  $x(t)$ 's nearest exemplar and correspondingly outputs the  $C_{near}$ .

**Function AKNN2** inputs  $x(t)$  and  $C_{near}$ , outputs  $x(t + \lambda)$ ,

$$(x(t + \lambda)) = AKNN2(x(t), C_{near}, k), \quad (33)$$

Calculate the similarity between  $x(t)$  and each point  $x_i$  in the same cluster  $C_{near}$  as

$$S_{im}(x(t), x_i) = -\|x(t) - x_i\|^2. \quad (34)$$

When  $S_{im}(x(t), x_j), j = 1, 2, \dots, k$  achieves the first  $k \max(S_{im})$  values,  $x_j, j = 1, 2, \dots, k$  are the nearest neighbors of  $x(t)$ . Thus the prediction  $x(t + \lambda)$  can be generated as

$$x(t + \lambda) = \frac{\sum_{j=1}^k x_j}{k}. \quad (35)$$

According to Eq. (35), the parameter  $k$  is a critical factor which can affect the results, thus find the optimal  $k$  value is a practical problem. In general, different specific historical data set has different optimal  $k$  value which is usually found by experiments. Therefore, in this paper, we find each optimal  $k$  value for each specific model by experiments, setting the original  $k$  value as  $k = 1$ , generating predictions based on nearest neighbors and increasing value with step length  $\lambda = 1$ . If  $k$  value continues to increase three times with no effect on improvement of prediction, we can stop and find the optimal  $k$  value makes the best performance of prediction. However, other methods should be studied in future work.

## 2.4 Three structural parameters of an EPAK financial prediction model

A specific EPAK financial prediction model can be constructed by three structural parameters: 1)  $\lambda$  is the prediction step length, 2)  $w$  is the sliding window width used for a forward rolling EMD, and 3)  $k$  means the nearest neighbors selected by the second layer of AKNN. Thus Eq. (7) which is a general definition of an EPAK model can be re-expressed as

$$R(t + \lambda) = AKNN\{AP[PCA(EMD(DR(t, w)))]\}, k\}. \quad (36)$$

### 3 A complex prediction model for stock market index synthesizing all sector indices predictions using EPAK as a kernel

A stock market index can reflect the overall trend of the stock market such as the Chinese benchmarked stock index CSI 300, the Standard and Poor's Composite Index and so on. It is generally calculated by Paasche Index as

$$I_r = \frac{CAV_r}{b} \times 1000, \quad (37)$$

where  $CAV_r$  and  $b$  indicate the adjusted market values of the constituent stock ( $CAV$ ) during the reporting period and on the base date, and  $b$  equals a constant that set as a divisor. Moreover,  $CAV_r$  can be calculated as [2]

$$CAV_r = \sum_{i=1}^n P_i \times AS_i, \quad (38)$$

where  $n$  is the number of total constituent stocks,  $P_i$  is the price of each constituent stock, and  $AS_i$  is the number of adjusted share capital of each constituent stock. Note according to Paasche Index, the circulation or volume of the constituent stock during the report period is set as a weight  $cw_i$ , thus  $AS_i$  can be calculated as

$$AS_i = TS \times cw_i, \quad (39)$$

where  $TS$  is the number of total share capital. Therefore, Eq. (37) can be rewritten as

$$I_r = \frac{1000}{d} \times \sum_{i=1}^n TS \times P_i \times cw_i, \quad (40)$$

In general, the constituent stocks of the stock market index are belongs to different industries, accordingly different sector indices can be calculated. Therefore, we can construct a complex prediction model for stock market index synthesizing all sector indices predictions. Applying an EPAK as a kernel, this complex model can be expressed as

$$MIP = \sum_{i=1}^N SIP_i \times sw_i, \quad (41)$$

where  $SIP_i$  is the prediction of each sector index generated by EPAK,  $sw_i$  is the a sector weight which sums the weights of the constituent stocks, and  $N$  is the number of total sector indices.

## 4 Empirical test and results

To test the effectiveness of an EPAK model for predicting a real data set, we do an empirical test on Chinese stock index and collect real historical data set from Wind and China Securities Index Co., LTD (China).

### 4.1 Performance metrics of a specific EPAK prediction model

For evaluating the performance of a prediction model, some metrics can be applied, generally including Mean Absolute Percentage Error (MAPE), Mean Absolute Difference (MAD), and

Root Mean Square Error (RMSE). However, we apply Hit Rate [14] that can reflect accuracy of the predicted direction as a metric in this paper. It is defined as

$$HitRate = \frac{\sum_{i=1}^n h_i}{n}, h_i = \begin{cases} 1, & R_i \times R_p > 0 \\ 0, & R_i \times R_p < 0 \end{cases}, \quad (42)$$

where  $n$  indicates the number of samples,  $R_i$  and  $R_p$  are the real and predicted values.

## 4.2 A specific EPAK model for predicting CSI 300

Model EPAK\_CSI300d1 is proposed to predict the  $t+1$  daily return of Chinese benchmarked stock index CSI 300. This model can be expressed as

$$R(t+1) = AKNN\{AP[PCA(EMD(CSI300d1\_DR(t, w)))]\}, k\}. \quad (43)$$

The historical CSI 300 price time series from 4<sup>th</sup> January 2006 to 29<sup>th</sup> December 2017 comprises 2917 trading days is used, with the earlier 80% part for in-sample training and the later 20% part for out-of-sample testing. In terms of hit rate, the performance of EPAK\_CSI300d1 is shown in Table 1 (Note: The level of IMFs is determined by the EMD decomposition of CSI 300), resulting a highest hit rate of 72.78% with  $w = 100$ ,  $n = 3$  and  $k = 1$ . Therefore, we can say that EPAK\_CSI300d1 can predict the  $t+1$  daily return of CSI 300 effectively.

Table 1: Hit rates of EPAK\_CSI300d1 for  $t+1$  daily return of CSI 300 prediction

Hit Rate%					
w	k=1	k=2	k=3	k=4	k=5
100	72.78	71.35	68.12	68.29	70.63
150	72.29	70.27	71.19	67.90	67.90
200	70.47	69.91	68.98	69.35	68.05
250	70.85	70.66	67.82	70.10	67.44
300	72.22	70.29	69.13	67.58	69.32
350	68.92	72.66	71.48	68.32	69.31

## 4.3 A specific complex model for predicting CSI 300 synthesizing ten sector indices predictions using EPAK as a kernel

The constituent stocks of CSI 300 are belongs to ten sectors, thus we construct ten specific EPAK models for predicting the  $t+1$  daily return of these ten sector indices as shown in Table 2 (Note: The data is collected from <http://www.csindex.com.cn/>). The historical data set for each sector index during the same period as CSI 300's is used. Table 3 shows the test results, where the highest hit rate of each model is selected and the best performance in terms of hit rate is 79.60% with the EPAK model predicting the  $t+1$  daily return of Telecom Svc index. According to Eq. (41), we construct a prediction model for predicting CSI 300 synthesizing ten sector indices predictions as

$$CSI300 = \sum_{i=1}^{10} R_i(t+1) \times sw_i, \quad (44)$$

where  $R_i(t+1)$  is the prediction of each sector index generated by the EPAK prediction model. This model test achieves a hit rate of 73.43% which is higher than the performance of

Table 2: Ten sector indices form CSI 300

Index	Weight (%)	Constituent Number
Energy	2.61	12
Materials	7.04	33
Industrials	13.31	62
Cons Disc	11.92	37
Cons Staples	7.97	11
Health Care	5.15	22
Financials	39.81	75
Info Technology	7.62	30
Telecom Svc	2	7
Utilities	2.57	11

Table 3: Hit rates of EPAK models for daily return of ten sector indices prediction

Index	Hit Rate (%)	Index	Hit Rate (%)
Energy	73.54	Health Care	74.55
Materials	77.58	Financials	69.49
Industrials	78.59	Info Technology	71.52
Cons Disc	78.58	Telecom Svc	79.60
Cons Staples	76.57	Utilities	73.54

EPAK\_CSI300d1. This result implies the complex prediction model for prediction the stock market index can improve the effectiveness of the prediction model with more comprehensive information as input.

## 5 Conclusion

This paper proposes a financial prediction model called EPAK, integrating a nonlinear feature extraction procedure and a two-layer KNN with AP for prediction. The nonlinear feature extraction procedure is adaptable and comprehensive for financial time series analysis and the new two-layer KNN with AP can tackle the main deficiencies of KNN and perform better. Applying the EPAK model as a kernel, we construct a complex model for stock market index prediction which synthesizes the predictions of each sector index of the stock market via weighted average to generate the prediction of the stock market index. Specific EPAK models for univariate time series are implemented for predicting the  $t + 1$  daily return of CSI300 and ten sector indices of CSI300, achieving a highest hit rate of 79.60% on Telecom Svc index prediction. Then the predictions of ten sector indices are synthesized via weighted average for generating the prediction of CSI 300, which performs better than the direct prediction of CSI 300.

As a prediction model for financial time series, EPAK comprises two key processes, feature extraction and modeling of prediction which are also the main factors for the performance of prediction. Thus in order to improve the prediction model, we can focus on these two procedures in our future work. For feature extraction, we can advance four aspects on: 1) taking more comprehensive information from different financial markets that interact with each other [7,24], 2) applying other effective nonlinear dimension reduction algorithm integrating other methods which should be more suitable for financial time series, 3) finding a similarity metric to take place of Euclidean measurement which is special for financial time series, 4) improving the prediction modeling, such as apply the Auto Encoder, Random Forest and so on.

## Bibliography

- [1] Beckers, B.; Herwartz, H.; Seidel, M. (2017); Risk forecasting in (T)GARCH models with uncorrelated dependent innovations, *Quantitative Finance*, 17(1), 121-137, 2017.
- [2] China Securities Index Co., LTD (China). CSI 300 index compilation scheme. Shanghai: China Securities Index Co., LTD (China); 2016.
- [3] Davidson, J.; Li, X. Y. (2016); Strict stationarity, persistence and volatility forecasting in ARCH process, *Journal of Empirical Finance*, 38, 534-547, 2016.
- [4] Dudoit, S.; Fridlyand, J. (2002); A prediction-based resampling method for estimating the number of clusters in a dataset, *Genome Biology*, 3(7),1-21, 2002.
- [5] Frey, B. J.; Dueck, D. (2007); Clustering by passing messages between data points, *Science*, 315, 972-976, 2007.
- [6] Huang, N. E.; Shen, Z.; Long, S. R. ; et al. (1998); The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society A:Mathematical, *Physical and Engineering Sciences*, 454, 903-995, 1998.
- [7] Hu, C.; Liu, X.; Pan, B.; et al. (2017); Asymmetric Impact of Oil Price Shock on Stock Market in China: A Combination Analysis Based on SVAR Model and NARDL Model, *Emerging Markets Finance and Trade*, 2017 (just-accepted).
- [8] Iabal, M.; Naveed, A. (2016); Forecasting inflation: Autoregressive integrated moving average model, *European Scientific Journal*, 12(1), 83-92, 2016.
- [9] Jaramillo, J.; Velasquez, J. D.; Franco, C. J. (2017); Research in financial time series forecasting with SVM: Contributions from literature, *IEEE Latin America Transactions*, 15(1),145-153, 2017.
- [10] Jena, P. R.; Majhi, R.; Majhi, B. (2015); Development and performance evaluation of a novel knowledge guided artificial neural network (KGANN) model for exchange rate prediction, *Journal of King Saud University-Computer and Information Sciences*, 27(4), 450-457, 2015.
- [11] Karl Pearson, F. R. S. (1901); On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, 2, 559-572, 1901.
- [12] Martinez, F.; Frias, M.; Perez, M. (2017); A methodology for applying k-nearest neighbor to time series forecasting, *Artificial Intelligence Review*, 1-19, 2017.
- [13] Mokoma, T. J.; Muroke, N. D.(2015); Is the South African exchange rate volatile? Application of the ARCH framework, *Risk Governance and Control: Financial Market & Institutions*, 5(1), 110-122, 2015.
- [14] Pan, H. P.; Haidar, I.; Kulkarni, S. (2009); Daily prediction of short-term trends of crude oil prices using neural networks exploiting multimarket dynamics, *Frontiers of Computer Science in China*, 3(2),177-191, 2009.
- [15] Pan, H. P.; Ma, Y.; Zhang, C. Z. (2017); FEPA-An integrated computational intelligence model for predicting financial time series, 2017 IEEE/SICE International Symposium on System Integration (SII 2017), 2017 Dec 11-14, Taiwan, China (Accepted).

- 
- [16] Pedro, C. S.; Pedro, H. M. (2017); Volatility forecasting via SVR-GARCH with mixture of Gaussian kernels, *Computational Management Science*, 14(2), 179-196, 2017.
- [17] Ravi, V. ; Pradeepkumar, D. Deb, ; K. (2017); Financial time series prediction using hybrids of chaos theory, multi-layer perceptron and multi-objective evolutionary algorithms, *Swarm and Evolutionary Computation*, 36, 136-149, 2017.
- [18] Rotshtein, A.; Pustylnik, L.; Giat, Y.(2016); Fuzzy logic and chaos theory in time series forecasting, *International Journal of Intelligent Systems*, 31(11), 1056-1071, 2016.
- [19] Sermpinis, G.; Stasinakis, C.; Theofilatos, K.; Karathanasopoulos, A. (2015); Modeling, forecasting and trading the EUR exchange rates with hybrid rolling genetic algorithms-Support vector regression forecast combinations, *Harvard Business Review*, 247(3), 831-846, 2015.
- [20] Tealab, A.; Hefny, H.; Badr, A. (2017); Forecasting of nonlinear time series using ANN, *Future Computing and Informatics Journal*, 2(1), 39-457, 2017.
- [21] Wang, J.; Wang, J. (2017); Forecasting stochastic neural network based on financial empirical mode decomposition, *Neural Networks*, 90, 8-20, 2017.
- [22] Wang, K.; Zhang, J.; Li, A.; et al. (2007); Adaptive affinity propagation clustering, *Acta Automatica Sinica*, 33(12), 1242-1246, 2007.
- [23] Wen, F.; Gong, X.; Cai, S. (2016); Forecasting the volatility of crude oil futures using HAR-type models with structural breaks, *Energy Economics*, 59, 400-413, 2016.
- [24] Wen, F.; Xiao, J.; Huang, C.; et al. (2018); Interaction between oil and US dollar exchange rate: nonlinear causality, time-varying influence and structural breaks in volatility, *Applied Economics*, 50(3), 319-334, 2018.
- [25] Yang, H. L.; Lin, H. C. (2017); Applying the hybrid model of EMD, PSR, and ELM to exchange rates forecasting, *Computational Economics*, 49(1), 99-116, 2017.
- [26] Zhang, C. Z.; Pan, H. P. (2015); A forecasting model based on forward rolling EMD techniques, *Technical Economics (a Chinese Journal)*, 34(5), 70-76, 2015.
- [27] Zhang, G. S.; Zhang, X. D.; Feng, H. Y. (2016); Forecasting financial time series using a methodology based on autoregressive integrated moving average and Taylor expansion, *Expert Systems*, 33( 5) , 501-516, 2016.
- [28] Zhang, N. N.; Lin, A. J.; Shang, P. J. (2017); Multidimensional k-nearest neighbor model based on EEMD for financial time series forecasting, *Physica A: Statistical Mechanics and Its Applications*, 477, 161-173, 2017.

# Outlook of Coordinated Transmission Control in 5G Networks for IoTs

U. Tariq, A. Aldaej

## Usman Tariq\*

Department of Information Systems  
College of Computer Science and Engineering  
Prince Sattam bin Abdulaziz University, Saudi Arabia  
\*Corresponding author: u.tariq@psau.edu.sa

## Abdulaziz Aldaej

Department of Information Systems  
College of Computer Science and Engineering  
Prince Sattam bin Abdulaziz University, Saudi Arabia  
a.aldaej@psau.edu.sa

**Abstract:** The sum of wireless nodes are forecasted to ascent sharply from about ten billion in year 2018 to twenty-five billion by 2020. Consequently, the data packet capacity response is anticipated to increase in future. The projected 5th generation (5G) of cellular grids is anticipated to be a blend of grid mechanisms with diverse magnitudes, transfer energy, backhaul networks and wireless access equipment. Although there are several fascinating complications comprised by the 5G context, we highlighted the problems of communication mechanism with the potential of Internet of Things (IoT). We offer a charter to discover required method constraints strategy that deal with the preeminent advances in terms of possible output for device-to-device inspired mobile grids. We explore the working and breakdown of immediate data and energy handover in IoT and assess individually about communication outage likelihood and energy outage possibility.

**Keywords:** wireless communication, receiving antennas, 5G, internet of things (IoT)

## 1 Introduction

To assess the payload communication response on the mobile grid, in entities of Gigabit/second/ km square, the method implemented in reference [5] initiates with identified inhabitants concentrations in year 2010 for dissimilar setting types [1] and supposing a ratio of 1percent per annum for upsurge in subscriptions [2], then analyzes and marks supplementary forecasting of demand to year 2025 and beyond.

In the present day, with the current long-term evolution-advanced criteria, the 3GPP provision's a distinct-hop transmission tool in which the wireless admittance association between the base station (BS) and operator is transmitted by only one spread station. With the support of multi-hop transmitter, the wireless association between the base station and operators can be stretched to more than two nodes to expand the handling and grid capability. Several nodes communicate payload to and from the resultant base station effects in the decline of route loss. Nevertheless, practicing a multi-node transmitter method entails additional wireless resources to spread data from end to end nodes. Additional intrusion is also formed due to a larger sum of synchronized communications in the system. Novel agility and broadcast controlling systems are thus significant for accomplishing an extraordinary quality of service, though growing the entire cellular grid's ability. Over the preceding years, cooperative cellular data exchange [3] have achieved noteworthy devotion in the research communal. The key notion is the attainment of spatial uniformity devoid of consuming as a criterion the presence of several transceivers in distinct stations. More precisely, the respective cellular device turns out to be a fragment of

an enormous disseminated group, sharing its distinct aerial to promote the transaction among devices, using either amplify-and-forward or decode-and-forward schemes. As an outcome, the concluding endpoint can collect several reproductions of the similar data, which can be pooled to advance the consistency of the communication. Consequently, disseminated collaborative activities can successfully transmit the wireless data, by possibly provisioning: (i) complex spatial range and output; (ii) lesser battery intake and condensed intrusion; and (iii) compliance to system settings.

## 1.1 The Internet-of-Things (IoT)

The Internet has discarded substantial barriers in linking inhabitants, organizations, and neighborhoods universally by focusing diverse areas of interest. Transceiver motes produce information, information generates understanding, and understanding impels accomplishment. As a result, establishing logic of the information that individual nodes produce does have the authentic significance for inhabitant. The creation of context aware cellular devices has created an ample recognition and acceptance by industry, user and common residents. At the moment, we observe soaring escalation in cellular traffic than traditional phone lines. Enhanced energy capacity, adaptation of smart devices and expansion of network for cellular functions will affect a steady downfall in the augmentation of personal computers. With the mounting implementation of context aware devices and group medium, society event tracking and consequently inhabitant activity information are available for further processing at an enlarged scale.

### Applications

The idea of the internet of things is not novel. To hand, numerous premature IoT samples are available in the arena of engineering mechanism, development device, and cellular networks. We are presently coming into contact with the IoT, where masses of innovative nodes are frequently being coupled to the network. Since this phenomenon augment abilities such as environment cognizance, amplified energy/battery handling, and power objectivity, and as new individuals and novel categories of data are associated, we will rapidly come into the Internet of Everything, where gears that were quiet will have an expression.

**Well-being and Robustness:** IEEE 802.15.4 based tools are permitting us to observe without difficulty the acute life-symbols of people with special needs. Formerly, the absence of proper feelers, shared with an unfeasible mesh of cables, has rigorously restricted healing applications. At the present, patients can be examined successfully for grave disorders such as slumber apnea, even though in couch napping. For observing serious constraints such as pulse frequency, even though a sick is experiencing vigorous bodily workout, the linkage of cables is unworkable. This has advanced where individuals can now practice IoT over body area based connected nodes such as handheld devices that not only observe interval and achieve messages but also screen the diabetic level in a user's blood devoid of the discomfort and distress of a spike puncture. The IEEE P2413 based applications have achieved milestones in medicine and health.

**Agile Inhabitants:** Prospect homes will have segments of linked nodes into a principal structure that can empower the user with a computing device to govern everything in functioning a smart-home. With a dash of a control, the handler can use device over any application in the connected environment such as safekeeping, temperature device, or smart theatre. Try on the homegrown multimedia system, forming spontaneously the seamless environment, acoustic arrangement, and others can all be functioned from a particular application.

**Industry:** As per CISCO, currently Manufacturing Network Grid is worth thirty two trillion dollars. By the year 2030, it could range to eight five trillion dollars of yield. At production

facility, the IoT benefits the enterprise to accumulate information about progressions going on daily bases. With the help of thousands of feeler devices on the assemblage stripe and the devices in each sequence, administrators can promptly discover the position of assembly. They are capable to exchange that statistic with collaborators in additional subdivisions. Devices at CISCO has stemmed in modernizations never previously comprehended. [4]

### **Smart Environment Architectures and Applications**

The alteration of the current system to the requirements of IoT unquestionably has a lot of benefits because of its natural capacity to collaborate with the current link node grid structure. The advancement of IoT which is built on presently adopted Internet Protocol, Uniform Resource Identifier and Domain Name Server tools, can be directed on the foundation of previously learned awareness and understanding, with the practice of the prevailing, normally adopted systems and methods.

IoTSyS [8] delivers an adaptation level to empower the incorporation of entities in the IoT design. It is dedicated on enabling data exchange for rooting nodes, which are exhausting IPv6. The emphasis on the adaptation framework is on improving the compatibility of nifty entities. The adaptation framework exploits the Internet Protocol v6 based Low-power Wireless Personal Area Networks (LoWPAN), Constrained Application Protocol (CoAP) and extensible Markup Language substitution over the internet service area to relate with transceiver nodes. The IoTSyS adaptation framework targets at administering an access theory for current mote structures, set up currently in structure computerization, a mass which can be installed openly on rooted 6LoWPAN nodes.

With transportable networks taking noteworthy depends on the collaboration of the discrete resource sourcing devices that form network associations, the operator controlling a node may adopt to the supportive assignation. Consequently a precise supporting module is the non-methodological operational support, which resides on current wireless systems and innovative standards, such as grid programming. Intermediary function software methodologies, also measured as link level cumulating methods, perform as arbitrator executor both on a cellular node and inside the mobility aware network.

### **Technical Challenges**

Even though the right to use delay presented by Long-Term Evolution is satisfactory for supreme cellular broadband users, it possibly will not be enough for delay critical use cases, such as traffic wellbeing, structural safety, or developing built-up Internet systems. To guarantee provision for such operation-oriented cellular systems, subsequent peers of wireless right to use must permit for latencies on the demand of one millisecond or less.

### **Configuration and Scalability**

To empower the anticipated enormous traffic escalation, supplementary band will have to be allocated to cellular wireless infrastructures. In subsequent group equipment, the motivation will be on guaranteeing additional band beneath 6.5 gigahertz. Nevertheless, to accomplish extended traffic difficulties and capability empowerment of the very wide-ranging broadcast bandwidths, required for multi-gigabit per second payload rates, succeeding wireless right to use will outspread the choice of process to rates above ten gigahertz.

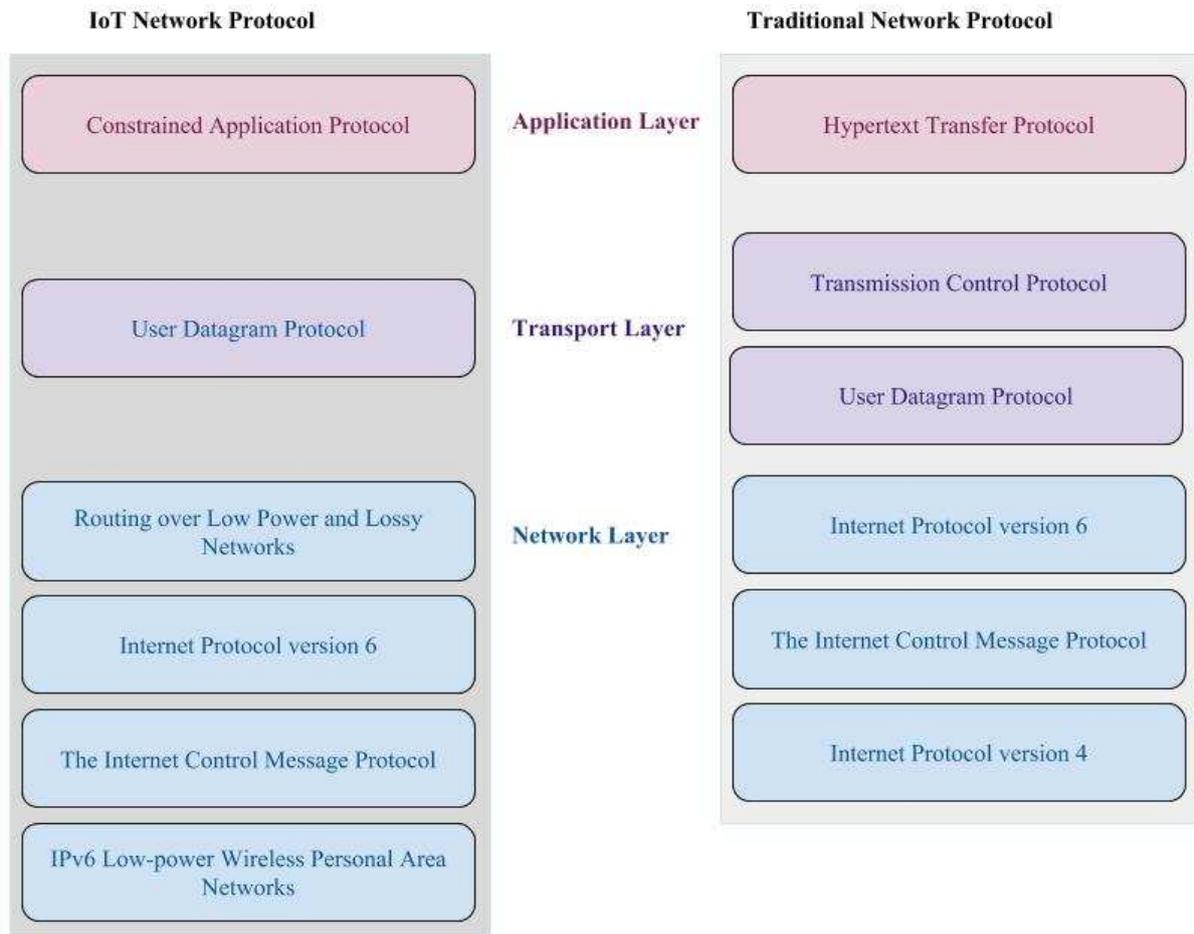


Figure 1: Data Transfer procedures of Internet of Things vs. customary system

### Mobility

Next generation feelers with several aerial essentials can advance exposure for extraordinary-data-ratio communications as well as considerably escalate global structure capability. Beamforming, where numerous feeler features are adopted to practice constricted rays, is a resourceful system for enhancing both information ratios and capability. Three-dimensional multiplexing, where broadcast features are utilized to deliver numerous information relays separately to solitary or supplementary terminals, is an alternative illustration of a vital multi-aerial procedure.



Figure 2: Beamforming for Long-Term Evolution and IEEE 802.16 standard devices

## 1.2 Network architecture

The evolution to a Schmoozed Civilization will empower to a substantial quantity of linked nodes, which spread trivial sums of information occasionally. Such nodes will regularly be unassuming and unnoticeably rooted into the inhabitant surroundings. This necessitates trivial wireless-component strategy and broadcast methods modernized to the undisturbed data exchange needs. Nodes have to be capable to function for longer sessions on minute batteries. It is required to accomplish nil-overhead data exchanges by streamlining linkage situations for nodes and granting rights to use network with nominal beckoning. Exploiting the node's snooze prospects can diminish battery utilization, resulting to extended operational uses.

The all-purpose consent on the necessities for 5G are:

- Procedure upsurge in zone capability with veneration to LTE-Advanced.
- One millisecond Round Trip Time delay.
- Procedure upsurge in battery productivity in relations of Joules per bit.
- Procedure decline in rate of setting out.
- Movement provision and continually-on the linkage of handlers that have extraordinary data requests.

### Protocols

Reflexive sub-interference conquest is illustrated as the indicator-power reduction executed by the route cost because of the physical parting among spread and collect feelers of the identical node. Classic reflexive sub-interference conquest methods consist of:

*Guiding sub-interference rout:* In this method, the key energy sections of the spread/collect feelers of a full duplex node have nominal connection, empowering the sub-interference to be moderately suppressed preceding to the collector's radio frequency visible-edge.

*Correspondent Withdrawal:* In correspondent withdrawal, the set of interval-field withdrawal systems such as preparation-centered procedures can be hired together by "distinct-response distinct-yield" and "numerous-response numerous-yield" oriented procedures, where the former may execute sub-interference destruction by misusing the three-dimensional diversity accomplished by the linked numerous spread and collect feelers.

*Feeler Parting and Sub-Interference Elimination:* Aggregating the route cost amid the transfer/collect feelers set up an effective method to weaken the sub-interference control, in this technique a sophisticated feeler parting indicates improved sub-interference suppression routine. While depending on feeler parting, the regular inaccessibility may also abuse the adjacent structures or the valuable presence of a protective plate, only if that harsh limitations forced on the node magnitude can be gratified.

In conclusion, the unusual design will also entail the improvement of an integrated mechanism and administration strategy (grid prototype and edge evolution) to streamline grid procedures through diverse systems. It is assumed that the design/technique should be capable to harmonize with traditional/throwback framework and should agree to the handlers horizontal Hegira.

### Necessities for operators' data pooling technique

Novel elastic multiplexing and accumulation techniques, for instance, built in elastic-network and elastic-inactive optical grid. Cardinal frequency handling empowered bandwidth-adaptable

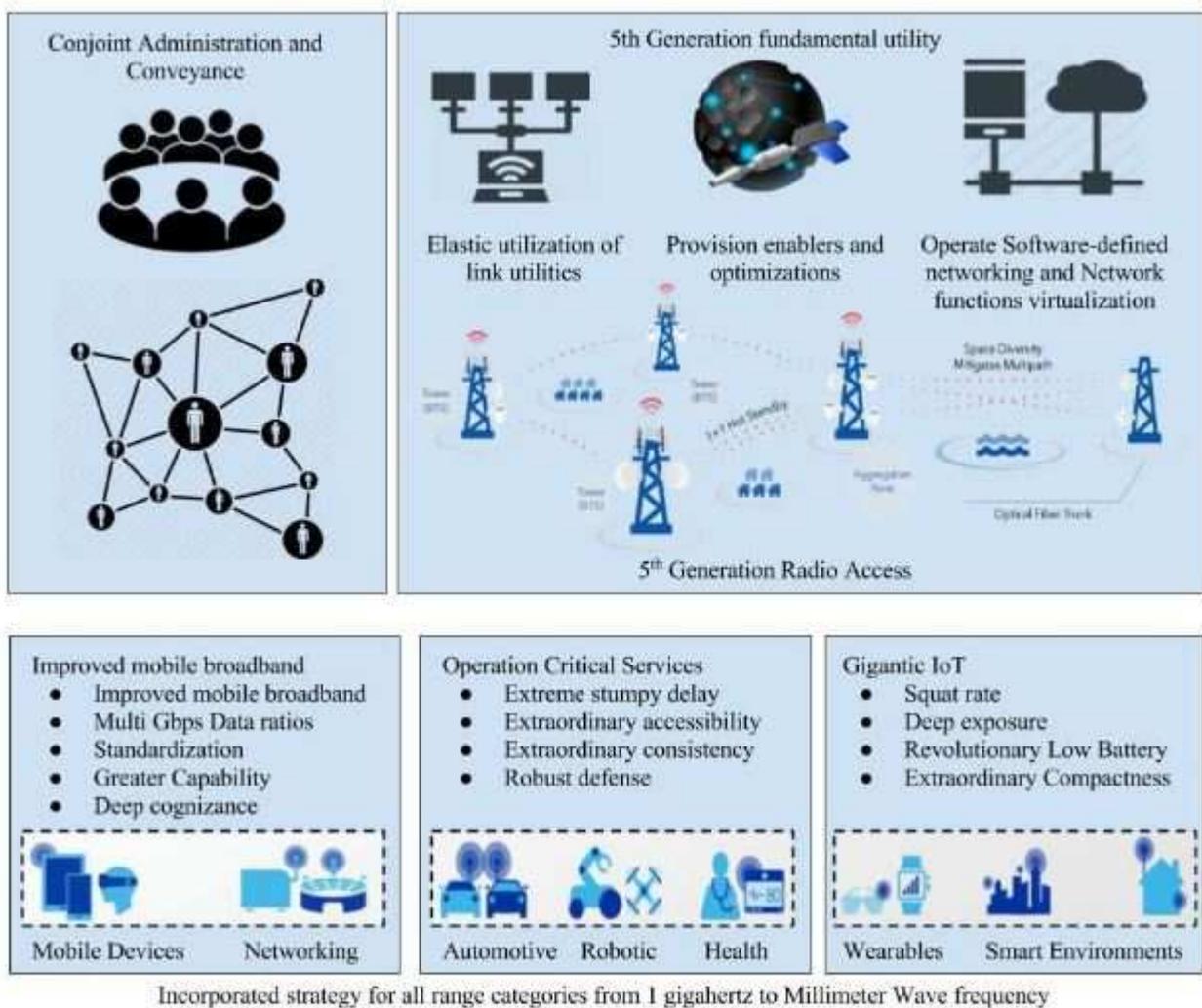


Figure 3: A conception of the general 5G design

optical signal detection devices will be measured to edge admission and grids, established on multi-transferor "OFDM" systems for varied inflection and multiplexing, will lead to improved represent-perceptivity detailed dataset at the control-focused level. This will outcome in the abundant formulation of the robust optical signal detection devices that can accomplish peak transmission provisions.

## 2 Related work

Innovative and interconnected environment necessitates not only over-the-top-fast integration for ingenious devices, and minicomputers, but also making accessible a gigantic sum of context-aware, resourceful, habitually stumpy-cost and squat-complexity nodes and making them linked even under very difficult settings. We believe IoT will be the ultimate system for 5th generation frameworks. What presently outlook as an obstacle of the IoT are incoherent devices. For instance, we partake radio-frequency identification, we use diminutive-array data relay methods, Ultra-wideband and Near-field communication, and this possibly will be a tricky in the forthcoming systems, especially in context of a smart metropolitan, where an integrated structure for

continuous node linking is essential. 5th generation networking is a worthy prospect to offer this integrated context.

5G will improve the signal route communications to guarantee broad scope for these nodes. For the up-interconnection communication, 5G can optionally provide the procedure for non-quadratic diverse connection methodologies for illustration such as random up-interconnection data relay from the battery-oriented IoT transceiver sensing nodes. 5G nodes will also provision route oriented data connections, achieved by the grid. This will tolerate communications to be spread in anticipation that uninterrupted coverage is possible, which permits these nodes to consume lesser control communications.

## 2.1 Multimeter wave channel

In the preceding few years, the 28 gigahertz, 38 gigahertz and 80 gigahertz rate of recurrence bands have converted as an essential component for millimeter waves cellular and adhoc data relays [6] [7] [9]. Lately, millimeter waves broadcast depth operations have been examined in enclosed and compact metropolitan open-air surroundings [10] [11]. Millimeter waves broadcasting-methods are likely to function in substantial guided communications and it should respond to the routine of electrically-navigated multi-portion feeler groups to justify for the degree surge in unrestricted space route loss. Expected Omni functional route loss in compact metropolitan millimeter wave's network is important for method policy and for approximating exposure and the ability of developing revolutionary extensive band cellular grids. Academics' practice interval graphs and three-dimensional sections to perfect the omni-indicator "committed information rate" and analogous joint angle-of-departure and angle of arrival control ranges, which have been adopted effectively in forming Millimeter wave frequencies [15].

## QODM

Multi-bearer variation in the practice of quadratic occurrence disunion multiplexing (QODM) has succeeded in telecommunications due to its extraordinary frequency spread ability with great bandwidth effectiveness and its strength with respect to multi-route dwindling and extended intervals [16]. Core purpose that QODM as an explicit set of multicarrier inflection has been broadly adopted for extraordinary-data-ratio broadcasts in interval-dispersive settings which has substantial gains over interval-field comparison.

QODM separates an extraordinary-ratio data flow into N corresponding flows, which are then spread by controlling N discrete transferors. In case, the creditor wants to be capable to detach signals passed by diverse transferors, they have to be squarish.

Frequency approximation is an additional vital fragment of a logical QODM system. With frequency approximation, QODM methods can use logical exposure to gain a three decibel "signal to noise" proportion improvement over variance discovery.

For QODM methods with several spread and beneficiary feelers for system ability or routine enhancement, frequency data is indispensable to range merging, intrusion dominance, and frequency discovery. In practice, the precision of frequency state data significantly impacts the general scheme routine.

## FBMC

A substitute to QODM has been greatly endorsed: Filter bank centered multicarrier (FBMC) [14], [13]. It depends on sharing the band into numerous orthogonal substitute-bands, moreover, FBMC put on a filtering method to every substitute carrier in divergence to QODM. A method adopting FBMC is much more appropriate to a prospective 5G structure. Whereas in principle,

FBMC has several agreeable integrant, applied method arrangements reduces most of idle channels. As a multicarrier system, FBMC can value from multi-feeler channels. "Multiple input, multiple output" systems can also be useful.

### 3 Requirements for 5G

The pervasiveness of cellular correspondence, attached with a normalization on internet protocol (IP) as the data procedure, the accessibility of squat-cost and prevailing computing devices, service and data in the cloud, and growing data bandwidths "anticipated and conveyed" are approaching to add an idea of 5G. Simple necessities are illustrated as follows:

- Provision for stationary, movable, wireless and satellite connection tools;
- An accessible and elastic device grid that can be modified to the necessities from various amenities and horizontal markets (e.g., grid sharing, system utility virtualization);
- Resource effectiveness for amenities extending from squat data IoT facilities to extraordinary bit ratio software facilities;
- Battery productivity and energy control accession;
- Subsidiary integration from a distant user equipment (UE) through a spread UE to the system, and provision link between subsidiary networks and uninterrupted networks.

#### 3.1 Perceptive wireless networks and platforms

5G and Cognitive Broadcasting are the two developing protocols/systems to encounter the substantial cellular data flow of forthcoming wireless grids. Current developments in cognitive broadcasting tools have stressed the requirement for innovative and upgradeable feeler systems with improved output. One improvement of feelers when functioned for 5G is the infinitesimal proportions as 5G is concentrated to practice the millimeter wave frequencies.

Cognitive broadcastings are fully programmable cellular nodes and has a widespread revision provision for accomplishing improved system and operational performance. System should be context aware and must vigorously achieve adaptation in the grid procedures, band operation systems, frequency access techniques and broadcast waveform used. Its prospect will currently be determined by the organized research institutes that are functioning on novel methodologies with physical testbed oriented experimentation of cognitive broadcasting grids

#### 3.2 Current standards and application scenarios

Making 5G a genuine cutting edge futuristic industry standard demands a profound familiarity of what it requires to motivate novel portable protocols from regulation to industry and in what way these nodes will relate with other connectivity systems such as Wi-Fi (IEEE 802.11ac) and Bluetooth (IEEE 802.15.1).

- Keeping in focus for the mainstream requirements, the subsequent set of 5G provision is gaining industry recognition.
- 1 to 10 gigabits per second links to edge points in the inhabitant environments;
- 1 millisecond completion (i.e. Transmitter-receiver-transmitter) round trip latency;
- Ten to thousands of coupled nodes;

- 1000x transmission capacity per entity zone;
- Requirement of 99.99 percent accessibility;
- A 90 percent decrease in system battery usage;
- Requirement of hundred percent coverage;
- Up to five year energy availability for stumpy battery oriented nodes such as IEEE 802.15.4 standardized devices

## 4 Prospective work

Link layer clarifications built on the IEEE 802.19 wireless concurrence operational set regulation have three simple necessities: (a) the system edge program need to be harmonious with the transmission capacity accumulation procedure; (b) devices with dualistic or additional grid edges must fit into the identical decisive field; and (c) associations must be standardized. In an anticipated system, the quality of service of cognizant broadcast regulator practices a cross layer methodology to robustly react to the variations in the scheme constraints. It practices Automatic Repeat Request to boost the wireless association devoid of initiating a regular timeout. Furthermore, it upholds the end-to-end exposition of Transmission Control Protocol (TCP) since it does not send TCP acknowledgement to the contributor before the TCP sector is acknowledged effectively at the destination.

### 4.1 Synchronized multipath transmission (SMT)

In SMT, there is no dissimilarity concerning the main and minor routes. All routes are alike, and are adopted separately in data broadcast. SMT utilizes the identical method of continuance numbers as Stream Control Transmission Protocol. SMT practices the round-robin procedure to drive the packets along the access routes. The distribution of payload to the aerial is supervised by the magnitude of the bottleneck window for every route. SMT transmits the data end to end to a route as soon as the overcrowding window for required route becomes accessible. When quite a lot of routes can be adopted for the broadcast, the route is nominated by the round-robin method. SMT plugs the transmission overcrowding window of every route before hopping on to the subsequent.

Agility between mobile terminals and WLAN lacks proportionality. There are two categories of mobile-WLAN horizontal context aware node data transfers: voluntary handover where a cellular node is enclosed by both wireless grid and agrees to handover form solitary grid to alternative, and obligatory handover where a cellular node is departing the cluster of one system while exposed by alternative grid. In the course of a WLAN to mobile grid required horizontal handover, "Stream Control Transmission Protocol" routine is influenced by: the sinking of sequential packets on the WLAN connection for the reason that of the interruption in the wireless spectrum coverage, denoted as handover anomaly, and arbitrary data drain over wireless connection due to noise and intrusion, denoted to as a fault failure.

We witnessed, regardless of the gains of exhausting multi-route broadcast, SMT can perform de-grouping, which lowers the routines, considering the basis that the routes in SMT may have diverse features in terms of transmission capacity and latency.

SMT suggestions the method "Postponed Acknowledgement (PAC) for SMT" to diminish ACK data transmissions. PAC empowers the SMT to postpone the ACK, even if the payloads reach as disarray. As the transporting of ACK is repeatedly observed, the receiver should diligently evaluate every ACK expected to distinguish missed payloads as swiftly as promising. In

Table 1: Payload rates, mandatory for data set transfer in an assumed time intermission

**Note.** Comparative 5G data rate: 15bps/Hz (uplink) and 30 bps/Hz (downlink) top spectral productivity

	<b>Five minutes</b>	<b>One hour</b>	<b>Ten hours</b>
100 gigabits	03 Gbps	240 Mbps	30Mbps
10 gigabits	300 Mbps	25 kbps	3.1 Mbps
01 gigabits	30 Mbps	2.5 kbps	300 kbps
100 megabits	03 Mbps	250 kbps	30 kbps

Table 2: Interval to completely recover from bottleneck loss and data transfer during recovery session

<b>Handover Speed</b>	<b>Repossession Time</b>	<b>Data transfer in the course of recovery</b>
500 Mbps	08 minutes	14 gigabits
01 Gbps	15 minutes	55 gigabits
10 Gbps	125 minutes	500 Gbps

every ACK, the recipient updates the transmitter of the sum of payload packets it has acknowledged since the distribution of the last notifications. This suggestion necessitates adjustments to be made to Ariel (in both source and fetcher).

**Support of device mobility between networks**

We recommend to embrace Backhaul systems form the transitional grids between the entree systems to which the users are associated and the central grid. Such backhaul systems are suitable and ever more vital, because the trivial-cell mechanism entails an outsized number of links with all entry points, as a substitute of fixing the networks on a lesser number of large aerials. We can assume that backhaul grids principally comprised of associations between "Digital Subscriber Line Access Multiplexer" and the central system. Normally these links use Gigabit Ethernet or "Multiprotocol Label Switching" mechanism. In both cases, the guidelines for Stream Control Transmission Protocol packet broadcast and rebroadcast can be illustrated as follows: (a) as a definite rule, every edge node should constantly spread original data via the main route; and (b) a source should attempt to re-spread a chunk (which is a Stream Control Transmission Protocol data component) to a substitute active endpoint ID over the alternate route if broadcast over the crucial route has been unsuccessful.

**Low Latency Communication**

Delay continually outlooks at the forefront each and every time when mobile transmission is taken into the consideration. Overlook QoS, bit inaccuracy ratio, accessibility, consistency, circuit rapidity, terrestrial analysis or cost-value of a grid, the latency linked with communicating data over a mobile network continually escalate to the top, getting a reasonable portion of responsiveness. But delay in a mobile grid can differ significantly and researchers must be cognizant of all the diverse schemes to handle with this coefficient. We have acknowledged multiple adjustments to a futuristic device grid that would be required:

- Diminish communication time interludes, e.g., decrease it to "90 microseconds", and squatter QODM representative periods empowering firm and well-organized data communication;

- Reform of PHY networks permitting primary frequency assessment;
- Practice of standard codes (e.g., for data networks) and segment ciphers (e.g., for device frequencies) which can initiate fast and consistent deciphering;
- Adaptation of extraordinary uniformity intensities enhancement of the consistency of frequency recognition and deciphering, as well as accessibility.

It is suggested to adopt convolutional ciphering in the "accelerative inaccuracy rectification" schema for progression computerization.

## 4.2 Payload-collection (synthesization and reproduction)

Payload-Collection (PC) is a renowned tool in cellular dissemination, particularly in battery-controlled systems. PC entails in relating accumulation methods to review the grid node communications that streams over the routes. By decreasing the volume of grid communication, the devices lessen the battery usage events (e.g. Communication, response, congestion impact, and eavesdropping). Nevertheless, Payload-Collection carries a data aggregation drawback, subsequently it losses the level of Information-Precision and escalates the transmission weakness in information loss malfunctions. Some procedures necessitate extraordinary transmission consistency and greater information precision. Accordingly, in procedures with high importance, rather than decreasing the grid communication, the system must repeat the packets in order to intensify the consistency and guarantee enhanced information correctness. Packet-Reproduction is a method generally adopted in latency and interference-forbearing systems. In these systems/grids, integration is not certain, so the packet-reproduction can rise the consistency of data distribution.

One of the prominent packet-reproduction method is projected by Spyropoulos et al. [12] and it is termed Spray-and-Wait. In this scheme, an individual payload-creator device can reproduce its packets up to a determined number of copies. When the payload-creator device encounters a legitimate neighbor device, it transmits a copy to that nearby connect node. Once it dispenses all information copies, the payload-creator delays for the receiver endorsement notifying that the packet has been acknowledged. In case this verification is not received, the payload-creator device delays for the instance when it encounters the receiver and handover itself the packet.

On the subject of transference, which is precisely a key prerequisite for outsized grid node deployment, the interrelated tasks rely on neighbor devices, which are selected as device custodians. The anticipated context is aimed to be entirely disseminated, the synchronization of the roles does not rely on a neighboring or crucial entity. As an alternative, the devices accomplish discrete role assessments and modify these choices using data collected from a nearby cluster/zone.

## 4.3 Experimental validation

A number of occurrence-oriented applications in enormous scale and extremely compact IoT settings are associated with metropolitan systems where a device is allied to an outsized device network, and every device is capable to identify a set of occurrences. For instance, in a dense device grid, the devices function seamlessly and can give-and-take data linked to heat/smoke/fire recognition, automobile mishap alarm, and disaster and salvage procedures.

The validation scenario has been simulated using the "Network Simulator 3.28" [12] which was modified to accommodate as per intended scenarios. The features engaged are:

The capacity feature in figure 5 is the constraint, which is adopted to regulate the communication burden of the system; comprehensive explanation of the communication capacity aspect

Table 3: Emulation parameters

Factors	Assessment
Packet Size	1024 bits
Medium Access Control Header	300 bits
Physical Layer specific Preamble and Header	256 bits
Acknowledgement	224 bits + Physical Layer Header
Channel Bit Rate	1024 bits
Slot Time	65 microseconds
Short Interferframe Space	30 microseconds
Distributed coordination function Interframe Space	112 microseconds
Acknowledgement Timeout	250 microseconds
Contention Window (CWmax)	1 Megabyte
Transmit Power	2 Walts
Receiving Power	1.8 Walts
Sensing Power	1 Walts

Table 4: Packet layout of Code Division Multiple Access based Internet of Things payload communication: Packet Transmission Code

Preface Code	Source User Equipment Identification	Packet Length	Data Size	Cyclic Redundancy Check
56 bits	05 bits	05 bits	0 - 14 bits	15 bits

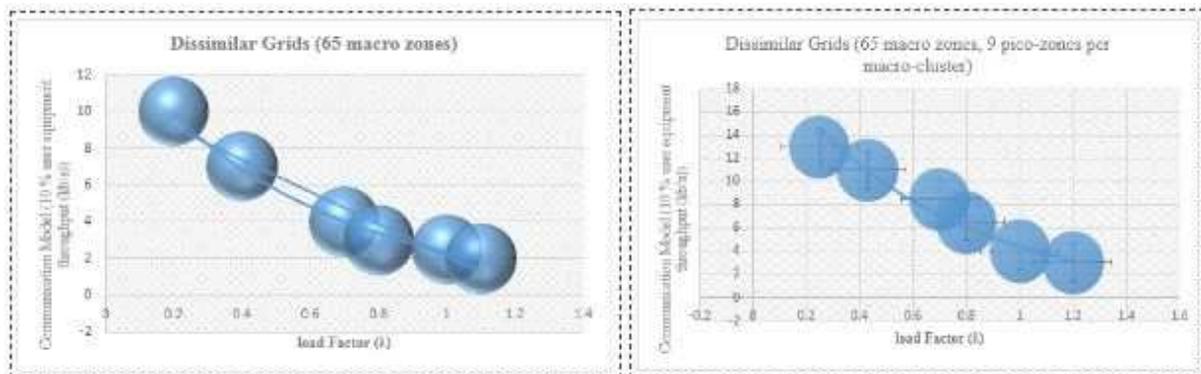


Figure 4: Illustration of Diverse and Similar grids for high payload communication

is in [17]. In dissimilar systems, the various pico-clusters establish supplementary zones, and the intersection of function and pico-clusters with diverse communication frequencies expands the interfering clusters. Thus, additional user equipment is authorized to use coordinated control in dissimilar systems. The normal cluster output in the dissimilar grid state appears to be worse than for the non-coordinated case, because diverse scheduler procedures were applied and evaluated. The performance outcomes cannot be accomplished by altering the scheduler limitations in distinct cluster processes.

## 5 Conclusion

5G systems/grids/nets should simplify the synthesis of numerous cellular and adhoc grids, which should tolerate for all standard applications and device of forthcoming smart cities. The practice of Millimeter wave band for 5G IoT operation is observed. In broad-spectrum, missing routes of diverse broadcast techniques is amplified in the millimeter wave spectrum. Although QODM is widely and consistently adopted in present communication schemes, inflection methods with enhanced spectral effectiveness are desirable to encounter the data ratio and grid capability necessities of application/hardware. Millimeter wave frequency coding is a dynamic study area that does not presently reveal a universally chosen resolution for 5G. The concept of multiple input, multiple output systems is systematically studied; MIMO operations were detached. Application of millimeter wave frequency is obligatory for well-organized IoT systems within 5G cellular grids, and respectively, the study determinations in the area stay to progressively increase.

## Acknowledgement

This project was supported by the Deanship of Scientific Research at Prince Sattam Bin Abdulaziz University, Alkharj, Kingdom of Saudi Arabia under the research project 2017/01/7170

## Bibliography

- [1] Bellanger, M.(2010); Physical layer for future broadband radio systems, *Radio and Wireless Symposium (RWS)*, 2010 IEEE, 436- 439, 2010.
- [2] Cisco Networks, *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013-2018*, March 2014.
- [3] Forbes, <http://www.forbes.com/sites/ptc/2014/07/01/how-the-internet-of-things-is-transforming-manufacturing/>, (accessed 18 November 2014).
- [4] Farhang-Boroujeny, B. (2011); OFDM Versus Filter Bank Multicarrier, *Signal Processing Magazine*, IEEE, 28(3), 92-112, 2011.
- [5] Li, B.; Zhou, S.; Stojanovic, M.; Freitag, L.; Willett, P. (2008); Multicarrier communication over underwater acoustic channels with nonuniform doppler shifts, *IEEE Journal of Oceanic Engineering*, 33(2), 198-209, 2008.
- [6] MacCartney Jr., G.R. et al. (2015); Indoor Office Wideband Millimeter-Wave Propagation Measurements and Channel Models at 28 and 73 GHz for Ultra-Dense 5G Wireless Networks, *IEEE Access*, 3, 2388-2424, 2015.

- 
- [7] MacCartney Jr., G.R. et al. (2015); Exploiting directionality for millimeter-wave wireless system improvement, *IEEE International Conference on Communications(ICC)*, June 2015.
- [8] Michahelles, F.; Ilic, A.; Kunze, K.; M. Kritzler, M.; Schneegass, S. (2017); IoT 2016, *IEEE Pervasive Computing*, 16(2), 87-89, 2017. doi: 10.1109/MPRV.2017.25
- [9] Nosratinia, A.; Hunter, T.E.; Hedayat, A.(2004); Cooperative Communications in Wireless Networks, *IEEE Communications Magazine*, 42(10), 74-80, 2004.
- [10] NS-3.28, <https://www.nsnam.org>
- [11] Palattella, M.R., Dohler, M., Grieco, A., Rizzo, G., Torsner, J., Engel, T. and Ladid, L., Internet of things in the 5G era: Enablers, architecture, and business models, *IEEE Journal on Selected Areas in Communications*, 34(3), 510-527.
- [12] Rappaport, T.S. (2013); Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!, *IEEE Access*, 1, 335-349, 2013.
- [13] Rappaport, T.S. (2015); Wideband Millimeter-Wave Propagation Measurements and Channel Models for Future Wireless Communication System Design, *IEEE Transactions on Communications*, 63(9), 3029-3056, 2015
- [14] Rappaport, T.S. et al., *Millimeter Wave Wireless Communications*, Pearson/Prentice Hall 2015.
- [15] Spyropoulos, T., Psounis, K., Raghavendra, C.S. (2005); Spray and wait: an efficient routing scheme for intermittently connected mobile networks, *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, ACM, 252-259, 2005.
- [16] Samimi, M.K.; Rappaport, T.S. (2015); Statistical Channel Model with Multi-Frequency and Arbitrary Antenna Beamwidth for Millimeter-Wave Outdoor Communications, *IEEE Global Telecommunications Conference (GLOBECOM)*, 1-7, 2015.
- [17] Vahid, S. Spectrum Floor for Future Mobile Communications Systems, *Internal 5GIC Study*, 2014

# Author index

Acho, L., 151  
Aldaej, A., 280

Bassam, A., 251  
Baulkani, S., 162

Caruntu, C.F., 151

Deng, Y., 205

Han, S.C., 175  
Hu, Y., 205

Ihler, A., 235

Jemilda, G., 162

Kang, B.H., 175  
Kim, D., 175

L, Y., 221  
Lee, S., 175  
Lin, S.J., 192  
Lin, Y., 175  
Liu, B., 205  
Liu, R.P., 192  
Liu, V., 221

Ma, Y., 235  
May Tzuc, O., 251  
Mendez-Monroy, P.E., 251

Ni, W., 192

Pan, B., 235  
Pan, H., 268  
Pujol, C., 151

Sanchez Dominguez, I., 251

Tang, L., 268  
Tariq, U., 280

Vargas, A.N., 151

Yu, J.G., 192  
Yuan, T., 221

Zhang, Z., 235