

**IJCSIS Vol. 13 No. 1, January 2015**  
**ISSN 1947-5500**

**International Journal of  
Computer Science  
& Information Security**

**© IJCSIS PUBLICATION 2015**



Cogprints

Google scholar



SciRate.com

CiteSeer<sup>x</sup> beta



DOAJ DIRECTORY OF OPEN ACCESS JOURNALS



ProQuest

# IJCSIS

ISSN (online): 1947-5500

Please consider to contribute to and/or forward to the appropriate groups the following opportunity to submit and publish original scientific results.

## CALL FOR PAPERS

### International Journal of Computer Science and Information Security (IJCSIS) January-December 2015 Issues

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.

See authors guide for manuscript preparation and submission guidelines.

Indexed by Google Scholar, DBLP, CiteSeerX, Directory for Open Access Journal (DOAJ), Bielefeld Academic Search Engine (BASE), SCIRUS, Scopus Database, Cornell University Library, ScientificCommons, ProQuest, EBSCO and more.

**Deadline:** see web site

**Notification:** see web site

**Revision:** see web site

**Publication:** see web site

Context-aware systems  
Networking technologies  
Security in network, systems, and applications  
Evolutionary computation  
Industrial systems  
Evolutionary computation  
Autonomic and autonomous systems  
Bio-technologies  
Knowledge data systems  
Mobile and distance education  
Intelligent techniques, logics and systems  
Knowledge processing  
Information technologies  
Internet and web technologies  
Digital information processing  
Cognitive science and knowledge

Agent-based systems  
Mobility and multimedia systems  
Systems performance  
Networking and telecommunications  
Software development and deployment  
Knowledge virtualization  
Systems and networks on the chip  
Knowledge for global defense  
Information Systems [IS]  
IPv6 Today - Technology and deployment  
Modeling  
Software Engineering  
Optimization  
Complexity  
Natural Language Processing  
Speech Synthesis  
Data Mining

For more topics, please see web site <https://sites.google.com/site/ijcsis/>



For more information, please visit the journal website (<https://sites.google.com/site/ijcsis/>)

## Editorial Message from Managing Editor

*The **International Journal of Computer Science and Information Security (IJCSIS)** promotes research publications which offer significant contribution to the computer science knowledge, and which are of high interest to a wide academic/research/practitioner audience. Coverage extends to several main-stream and state of the art branches of computer science, security and related information technology. As a scholarly open access peer-reviewed journal, IJCSIS mission is to provide an outlet for quality research & academic publications. It aims to promote universal access with equal opportunities for international scientific community; to scientific knowledge, and the creation, and dissemination of scientific and technical information.*

*IJCSIS archives all publications in major academic/scientific databases. Indexed by the following International agencies and institutions: Google Scholar, Bielefeld Academic Search Engine (BASE), CiteSeerX, SCIRUS, Cornell's University Library EI, Scopus, DBLP, DOI, ProQuest, EBSCO. Moreover, Google Scholar reported increased in number cited papers published in IJCSIS (**No. of Cited Papers: 524, No. of Citations: 1106, Years: 5**). Abstracting/indexing/reviewing process, editorial board and other important information are available online on homepage. By supporting the Open Access policy of distribution of published manuscripts, this journal ensures "free availability on the public Internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of [published] articles".*

*IJCSIS editorial board, consisting of international experts, guarantees a rigorous peer-reviewing process. We look forward to your collaboration. For further questions please do not hesitate to contact us at [ijcsiseditor@gmail.com](mailto:ijcsiseditor@gmail.com).*

*A complete list of journals can be found at:*

<http://sites.google.com/site/ijcsis/>

IJCSIS Vol. 13, No. 1, January 2015 Edition

ISSN 1947-5500 © IJCSIS, USA.

*Journal Indexed by (among others):*



## IJCSIS EDITORIAL BOARD

**Dr. Yong Li**

School of Electronic and Information Engineering, Beijing Jiaotong University,  
P. R. China

**Prof. Hamid Reza Naji**

Department of Computer Engineering, Shahid Beheshti University, Tehran, Iran

**Dr. Sanjay Jasola**

Professor and Dean, School of Information and Communication Technology,  
Gautam Buddha University

**Dr Riktesh Srivastava**

Assistant Professor, Information Systems, Skyline University College, University  
City of Sharjah, Sharjah, PO 1797, UAE

**Dr. Siddhivinayak Kulkarni**

University of Ballarat, Ballarat, Victoria, Australia

**Professor (Dr) Mokhtar Beldjehem**

Sainte-Anne University, Halifax, NS, Canada

**Dr. Alex Pappachen James (Research Fellow)**

Queensland Micro-nanotechnology center, Griffith University, Australia

**Dr. T. C. Manjunath**

HKBK College of Engg., Bangalore, India.

**Prof. Elboukhari Mohamed**

Department of Computer Science,  
University Mohammed First, Oujda, Morocco

# TABLE OF CONTENTS

## **1. Paper 31121405: Image Steganography Using Least Significant Bits (LSBs): A Novel Algorithm (pp. 1-5)**

*Abdul-Gabbar Tarish Al-Tamimi, Abdulmalek Abduljabbar Alqobaty  
Computer Science Department, Faculty of Applied Sciences, University of Taiz, Taiz, Yemen*

*Abstract* — Steganography plays a major role in data communication security. It focuses on hiding the fact that communication is taking place by hiding information in other information. The least significant bit (LSB) based approach is a popular type of steganographic algorithms in the spatial domain. The advantage of LSB is its simplicity to embed the bits of the message directly into the LSB plane of cover-image and many applications use this method. Images are the most popular cover objects used for Steganography. In this paper, we introduce a novel secure algorithm for color image steganography using least significant bits (LSBs). The security of our algorithm comes from using a key (the stego key) which is an array of 32 integers chosen randomly from  $Z_{24}$ . The first pixel position, where we start hiding process, is computed based on this stego key.

*Index Terms*—*information hiding, information security, steganography, LSB image steganography.*

## **2. Paper 31121409: Survey on Awareness of Privacy Issues in Ubiquitous Environment (pp. 6-10)**

*Huma Tabassum, Sameena Javaid, Humera Farooq  
Department of Computer Sciences, Bahria University Karachi, Pakistan*

*Abstract* — It is in human nature to keep certain details intimate. This confidential information, if made public, can result in annoyance, humiliation, and even devastating loss. The latter one holds true for individuals as well as organizations. In the case of ubiquitous environment, system adapts to the context of user's actions and intentions and the collection of data becomes pervasive. Thus, privacy turns out to be a vital aspect. This study aims to determine privacy awareness among people in ubiquitous environment. For this purpose, a survey was conducted. This survey was based on a questionnaire. The results show that people consider themselves quite aware, but, more concerned with privacy issues in ubiquitous environment. Also, a significant number of people admit to not taking privacy measures. The analysis was carried out on the basis of current status of participants in the university, namely, Undergraduate students, Graduate students, and Faculty members.

*Keywords*-*ubiquitous environment; privacy; privacy awareness.*

## **3. Paper 31121412: Efficient Support Vector Machines for Spam Detection: A Survey (pp. 11-28)**

*Zahra S. Torabi, Faculty of Computer Engineering, Islamic Azad University of Najafabad, Isfahan, Iran  
Mohammad H. Nadimi-Shahraki, Faculty of Computer Engineering, Islamic Azad University of Najafabad  
Isfahan, Iran  
Akbar Nabiollahi, Faculty of Computer Engineering, Islamic Azad University of Najafabad, Isfahan, Iran*

*Abstract* — Nowadays, the increase volume of spam has been annoying for the internet users. Spam is commonly defined as unsolicited email messages, and the goal of spam detection is to distinguish between spam and legitimate email messages. Most of the spam can contain viruses, Trojan horses or other harmful software that may lead to failures in computers and networks, consumes network bandwidth and storage space and slows down email servers. In addition it provides a medium for distributing harmful code and/or offensive content and there is not any complete solution for this problem, then the necessity of effective spam filters increase. In the recent years, the usability of machine learning techniques for automatic filtering of spam can be seen. Support Vector Machines

(SVM) is a powerful, state-of-the-art algorithm in machine learning that is a good option to classify spam from email. In this article, we consider the evaluation criterions of SVM for spam detection and filtering.

*Keywords- support vector machines (SVM); spam detection; classification; spam filtering; machine learning;*

#### **4. Paper 31121414: A Proposed Approach for Monitoring Quality of Web Services Using Service Level Agreement (pp. 29-36)**

*Nagy Ramadan Darwish, Department of Computer and Information Sciences, Institute of Statistical Studies and Research, Cairo University, Cairo, Egypt*

*Rabab Emad Mohamed, Department of Computer and Information Sciences, Institute of Statistical Studies and Research, Cairo University, Cairo, Egypt*

*Doaa Hany Elsayed, Department of Computer and Information Sciences, Institute of Statistical Studies and Research, Cairo University, Cairo, Egypt*

*Abstract* — Web service technology has gained more important role in developing distributed applications and systems on the Internet. Rapid growth of published web services makes their discovery more and more difficult. Nowadays, most of web service providers sign Services Level Agreement (SLA) contracts with their clients in order to guarantee the offered functionality of their services. This paper proposes an approach to monitor Quality of Services (QoS) in web service according to Service Level Objectives (SLO) in SLA. Monitoring procedures are introduced to check variations in the pre-agreed metric values of SLAs. Then, the deviation between the actual quality and the acceptable quality level can be identified and analyzed. Finally, the weaknesses of the web service practices can be discovered and solved.

*Keywords: Web Services, Services Level Agreement, Quality of Services, Web Service Level Agreement, Services Level Agreement Metric.*

#### **5. Paper 31121415: Towards a Fuzzy based Framework for Effort Estimation in Agile Software Development (pp. 37-45)**

*Atef Tayh Raslan, Department of Computer and Information Sciences, Institute of Statistical Studies and Research, Cairo University, Egypt*

*Nagy Ramadan Darwish, Department of Computer and Information Sciences, Institute of Statistical Studies and Research, Cairo University, Cairo, Egypt*

*Hesham Ahmed Hefny, Department of Computer and Information Sciences, Institute of Statistical Studies and Research, Cairo University, Egypt*

*Abstract* — Effort estimation in the domain of software development is a process of forecasting the amount of effort expressed in persons/month required to develop software. Most of the existing effort estimation techniques are suitable for traditional software projects. The nature of agile software projects is different from traditional software projects; therefore using the traditional effort estimation techniques can produce inaccurate estimation. Agile software projects require an innovated effort estimation framework to help in producing accurate estimation. The main focus of this paper is the utilization of fuzzy logic in improving the effort estimation accuracy using the user stories by characterizing inputs parameters using trapezoidal membership functions. In this paper, the researchers proposed a framework based on the fuzzy logic which receives fuzzy input parameters of Story Points (SP), Implementation Level Factor (ILF), FRiction factors (FR), and Dynamic Forces (DF) to be processed in many successive steps to produce in final the effort estimation. The researchers designed the proposed framework using MATLAB to make it ready for later experiments using real data sets.

*Keywords: Agile software development, Effort estimation, story points, fuzzy logic*

**6. Paper 31121420: A Novel Architecture for Improving Performance under Virtualized Environments (pp. 46-52)**

*A. P. Nirmala, Research Scholar, Karpagam University, Coimbatore & Assistant Professor, New Horizon College of Engg., Bangalore, India*

*Dr. R. Sridaran, Dean, Faculty of Computer Applications, Marwadi Education Foundation's Group of Institutions, Rajkot, India*

*Abstract* — Even though virtualization provides a lot of advantages in cloud computing, it does not provide effective performance isolation between the virtualization machines. In other words, the performance may get affected due to the interferences caused by co-virtual machines. This can be achieved by the proper management of resource allocations between the Virtual Machines running simultaneously. This paper aims at providing a proposed novel architecture that is based on Fast Genetic Kmeans++ algorithm and test results show positive improvements in terms of performance improvements over a similar existing approach.

*Keywords-* Virtualization, Performance, Performance Interference, Scheduling Algorithm, Throughput

**7. Paper 31121423: The Comprehensive Review on JDL Model in Data Fusion Networks: Techniques and Methods (pp. 53-60)**

*Ehsan Azimirad, PHD Student, Electrical and Computer Engineering Department, Hakim Sabzevari University Sabzevar, Iran*

*Javad Haddadnia, Associate Professor, Electrical and Computer Engineering Department, Hakim Sabzevari University Sabzevar, Iran*

*Abstract* — This paper provides the comprehensive review on JDL model in multi- sensor data fusion networks and its techniques and methods. Data fusion methods vary greatly depending on the type of problem and the surface to be integrated data. Two main motivations exist for using multiple sensors and combine the results of them. 1) Reduce errors and uncertainty in the measurements, 2) The use of multiple sensors to achieve a better estimate. With the data fusion of multiple databases, or multiple sensors and multiple natures, reduced uncertainty or ambiguity, and reduce complexities. The variety of models and architectures has been provided by researchers to combine sensor data for military and civilian applications. In this paper, we first defined the concept model, architecture and framework based on JDL model, and then states its techniques and methods.

*Keywords-* component; data fusion techniques; JDL model; data fusion models; data fusion architectures

# Image Steganography Using Least Significant Bits (LSBs): A Novel Algorithm

Abdul-Gabbar Tarish Al-Tamimi <sup>1</sup>, Abdulmalek Abduljabbar Alqobaty <sup>2</sup>

<sup>1,2</sup> Computer Science Department, Faculty of Applied Sciences, University of Taiz,  
Taiz, Yemen

<sup>1</sup> a.tarish@hotmail.com  
<sup>2</sup> qobaty@yahoo.com

**Abstract**—Steganography plays a major role in data communication security. It focuses on hiding the fact that communication is taking place by hiding information in other information. The least significant bit (LSB) based approach is a popular type of steganographic algorithms in the spatial domain. The advantage of LSB is its simplicity to embed the bits of the message directly into the LSB plane of cover-image and many applications use this method. Images are the most popular cover objects used for Steganography. In this paper, we introduce a novel secure algorithm for color image steganography using least significant bits (LSBs). The security of our algorithm comes from using a key (the stego key) which is an array of 32 integers chosen randomly from  $Z_{24}$ . The first pixel position, where we start hiding process, is computed based on this stego key.

**Index Terms**—information hiding, information security, steganography, LSB image steganography.

## I. INTRODUCTION

Steganography is the art and science of writing hidden messages in such a way that no one apart from the intended recipient knows of the existence of the message [1]. In contrast to cryptography, where the enemy is allowed to detect, intercept and modify messages without being able to violate certain security premises guaranteed by a cryptosystem, the goal of steganography is to hide messages inside other innocent messages in a way that does not allow any enemy to even detect that there is a second message present [2], [3], [4].

Among the methods of steganography, the most common thing is to use images for steganography. This is called image steganography. In this image hiding method, the pixels of images are changed in order to hide the secret data so as not to be visible to users, and the changes applied in the image are not tangible. The image used to camouflage the secret

data is called the cover-image while the cover-image with the secret data embedded in it is called the stego-image. Image steganographic techniques can be divided into two groups [5]: the Spatial Domain technique group, and the Transform Domain technique group. The Spatial domain technique embeds information in the intensity of the pixels directly, while the Transform domain technique embeds information in frequency domain of previously transformed image. Our proposed algorithm presented in this paper is a kind of the spatial domain techniques.

## II. RELATED WORK

The Least Significant Bit (LSB) steganography is one such technique in which least significant bit of a pixel (or a sub-pixel, for colored images) is replaced with data bit. In this method the least significant bits of some or all of the bytes inside an image is replaced with bits of the secret message. The LSB embedding approach is the basis of many techniques that hide messages within multimedia carrier data. LSB embedding is one of the most important steganographic techniques in use today [2].

Gandharba et al. in [6] proposed an approach to RGB channel based steganography technique which uses RSA algorithm for encryption and decryption. In an RGB image, each pixel (24 bits) is having R channel of 8 bits, G channel of 8 bits and B channel of 8 bits. The image is divided into 8 blocks and the cipher text is divided into 8 blocks. One cipher block is allocated to be embedded in only one image block by a user defined sub key. Out of the three channels in each pixel of the image one is used as the indicator channel. The indicator channel for the different blocks is not the same. The other two channels (called data channels) are used

for hiding cipher text bits in 4 least significant bit (LSB) locations. In a data channel 4 bits of cipher text can be embedded if after embedding the change in pixel value is less than or equal to 7. The two LSBs of indicator will tell whether the cipher text is embedded in only one data channel or in both data channels, so that retrieving can be done accordingly at the receiver. But pixel indicator techniques had a drawback that they treated all Red, Green, Blue components equally but in actual the contribution of all Red, Green, Blue components is not same for visual perception. So component based approaches were introduced.

Gutub et al., [7] and [8] introduced Pixel Indicator Techniques. They merged random pixel manipulation method and the stegokey to propose a technique, which uses the least two significant bits of one of the channels to indicate existence of data in the other two channels. He further targets high capacity in RGB image based steganography by introducing the concept of storing variable number of bits in each channel (R, G or B) of pixel based on the actual color values of that pixel i.e. lower color component stores higher number of bits. He also proposed Triple-A concealment technique method to hide digital data inside image-based medium. The algorithm adds more randomization by using two different seeds generated from a user-chosen key in order to select the component(s) used to hide the secret bits as well as the number of the bits used inside the RGB image component. The randomization adds more security especially if an active encryption technique is used such as AES. The capacity ratio is increased above SCC and pixel indicator scheme. Triple-A has a capacity ratio of 14% and can be increased if more number of bit is used inside the component(s).

Imran et al., in [9] proposed Component based Steganography for Color images. They integrated NPI, modified Least Significant Bits technique for data embedding and uses the green component of the image as it is less sensitive to human eye and thus it is totally impossible for human eye to predict whether the image is encrypted or not.

Chang et al, in [10] proposed a large payload data embedding method for color images. The proposed method modifies the blue value of the color pixel in order to imply the secret data because the blue value is an insensitive color to human eyes. Furthermore, the number of secret bits that can be embedded into

a cover pixel is dynamic and can be applied to both RGB and YUV color systems.

Roque et al, in [11] introduced a novel Steganography algorithm based on the spatial domain: Selected Least Significant Bits (SLSB). It works with the least significant bits of one of the pixel color components in the image and changes them according to the messages bits to hide. The rest of bits in the pixel color component selected are also changed in order get the nearest color to the original one in the scale of colors. This new method has been compared with others that work in the spatial domain and the great difference is the fact that the LSBs bits of every pixel color component are not used to embed the message, just those from pixel color component selected.

Ker et al. in [12] introduced the LSB matching scheme. LSB matching modifies the LSBs of the cover image for data hiding, but it does not simply replace the LSBs of the cover image as LSB replacement does. On the other hand, if one secret bit does not match the LSB of the cover image, then another one will be randomly added or subtracted from the cover pixel value. A revised version of LSB matching is proposed in [13] which greatly improve it by lowering the expected number of modifications per pixel (ENMPP), from 0.5 to 0.375. Therefore, the histogram affected by the scheme is less significant. In a generalization of LSB matching (G-LSB-M), sum and difference covering set of finite cyclic group were used to further reduce ENMPP and providing better security [14].

### III. THE PROPOSED ALGORITHM

The proposed algorithm consists of two sub-algorithms: the embedding algorithm and the extracting algorithm.

#### A. The Embedding Algorithm

In this section, we introduce the proposed embedding algorithm (Algorithm 1). The first step in embedding process is the stego-key generation step. The stego-key is an array of 32 integers randomly selected from  $\mathbb{Z}_{24} = \{0, 1, \dots, 23\}$ . Then first pixel position - where the embedding process starts - is computed. The embedding of the secret message is performed three characters from message in eight pixels (24 sub-pixels) from cover at a

time. The block of 24 bits (for the three characters) is circularly right shifted by the amount of bits corresponding the value of stego-key. Hence, the order of bits in `mblock` is randomly changed from one step (iteration) to the other (in one step, three characters are embedded in eight pixels).

```
Stego-Key: The key is an array  $K[32]$  of integers in the set  $\{0, 1, \dots, 23\}$  which are generated randomly.  
Input: The message to be hidden, the cover image of  $m \times n$  resolution, the hiding key  $K[32]$ .  
Output: Stego image.  
begin  
   $sum \leftarrow \sum_{i=0}^{31} K[i];$   
   $row \leftarrow sum^{sum} \bmod m;$   
   $column \leftarrow sum^{sum} \bmod n;$   
  First pixel position  $\leftarrow (row, column);$   
   $j \leftarrow 0;$   
  mblock[24] is a 24-bit message block;  
  str[3] is an array of three characters;  
  while !EOM do  
    /* EOM = End Of Message */  
    str  $\leftarrow$  read next three characters;  
    mblock  $\leftarrow$  convert str to binary;  
    RS = K[j](mblock);  
    /* RS = K[j] is a circular right shift function by K[j] bits */  
    replace the LSBs of the 24 sub-pixels of the next eight pixels by the corresponding bits in mblock;  
     $j \leftarrow (j + 1) \bmod 32;$   
  end  
end
```

**Algorithm 1:** The Embedding Algorithms

### B. The Extracting Algorithm

In this section, we introduce the extracting algorithm (Algorithm 2). The stego-key used for embedding is also used for extracting. Compute the first pixel position - where the extraction process starts - the same way as in the embedding algorithm. Then extract the least significant bits from 24 sub-pixels of next eight pixels and circularly shift them left by the amount of bits equals to the corresponding stego-key value. Continue until

the complete embedded message is extracted. The length of embedded message may be part of the stego-key, hence, the stego-key is the 32-integers array plus the length of the embedded message.

```
Input: The extraction key  $K[32]$  which is the same as the hiding key and the Stego-Image of  $m \times n$  resolution.  
Output: The hidden message.  
begin  
  compute the first pixel position as in the hiding algorithm;  
   $j \leftarrow 0;$   
  mblock[24] is a 24-bit message block;  
  str[3] is an array of three characters;  
  while not end of extracted message do  
    mblock  $\leftarrow$  read the 24 LSBs of the next 24 sub-pixels in the next eight pixels;  
    LS = K[j](mblock);  
    /* LS = K[j] is a circular shift left by K[j] bits */  
    str  $\leftarrow$  convert mblock to three characters;  
     $j \leftarrow (j + 1) \bmod 32;$   
  end  
end
```

**Algorithm 2:** The Extraction Algorithm

## IV. SECURITY ANALYSIS

Several steganographic algorithms have been proposed for embedding data in digital images as cover media.

The three most required evaluation criteria for good steganographic techniques are Robustness, Imperceptibility and Capacity [3]. The cover image is colored, and hence each pixel consists of three sub-pixels which are **red**, **green**, and **blue**.

### A. Robustness

The strength of our proposed algorithm depends on the hiding key which is an array of 32 integers in the set  $\{0, 1, \dots, 23\}$  which forms the first level of security. The hiding key is randomly constructed. Each element can be selected by 24 methods, and hence the total number of possible keys (the key

space) is  $24^{32} \simeq 1.5 \times 10^{44}$ , hence a huge number of possible keys. It is impractical for an attacker to exhaustively search this number of keys. A computer that works at  $10^6$  extractions/ $\mu s$  takes  $2.4 \times 10^{24}$  years to exhaustively search a half of the number of keys.

The first pixel position is random since it is computed using random values. This is the second level of security to our algorithm. The third level of security is that, each block of 24 bits (three characters) of the message is subjected to a transposition operation through circulating shift right by the corresponding key element. This transposition adds a complexity to steganalysis process. As a result, our proposed algorithm is secure (and hence robust) enough.

### B. Imperceptibility

Techniques that can be used to evaluate the imperceptibility of steganographic systems are different from one system to another depending on the type of cover file used for information hiding [15]. Since only the least significant bit of a sub-pixel is altered, it is visually imperceptible by human. The mean square error (MSE) between the cover image and stego-image and peak signal to noise ratio (PSNR) are used here as the measuring parameters for the amount of imperceptibility. MSE is defined as follows:

$$MSE = \frac{1}{mn} \left[ \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (C_{i,j} - S_{i,j})^2 \right]$$

where  $m, n$  are the number of rows and columns of the cover/stego image respectively.  $C_{i,j}$  is the pixel at row  $i$  and column  $j$  of cover image, and  $S_{i,j}$  is the pixel at row  $i$  and column  $j$  of stego image. The MSE is separately calculated for red sub-pixels ( $MSE_R$ ), green sub-pixels ( $MSE_G$ ), and blue sub-pixels ( $MSE_B$ ). The final MSE is  $MSE = (MSE_R + MSE_G + MSE_B)/3$ . The PSNR of the cover image is defined as follows:

$$PSNR = 10 \cdot \log_{10}(255^2/MSE)$$

The MSE and PSNR for stego image in Figure 1 were computed with MATLAB, Table I shows the values of MSE and PSNR at different sizes of the secret message which all show high imperceptibility.

### C. Capacity

The embedding capacity is the maximum number of bits that can be embedded in a given cover file [15]. The capacity of our proposed algorithm is 3 bits per pixel. The higher size the cover, the higher embedding capacity.

## V. IMPLEMENTATION RESULTS

Both embedding and extracting algorithms are implemented using C# programming language, see Figures 1 and 2. The evaluation of  $row \leftarrow sum^{sum} \bmod m$  and  $column \leftarrow sum^{sum} \bmod n$  is performed using the SQUARE-AND-MULTIPLY algorithm [4].



Fig. 1. The Embedding Algorithm

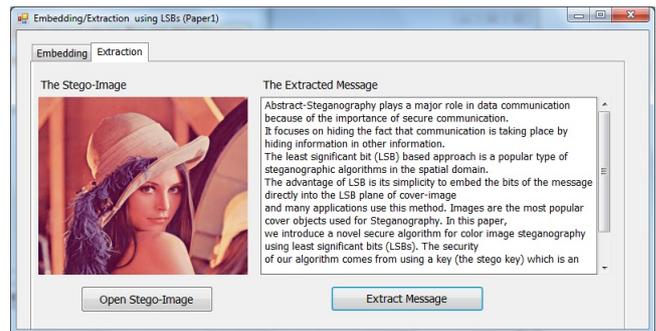


Fig. 2. The Extracting Algorithm

The following table shows some values for MSE and PSNR at different sizes of the message to be hidden.

## VI. CONCLUSION

In this paper, we proposed a novel algorithm for color image steganography using Least Significant

TABLE I  
MSES AND PSNRs AT DIFFERENT MESSAGE SIZES

Message Size	MSE	PSNR (dB)
915 B	$4.5 \times 10^{-3}$	71
1.78 KB	$9.04 \times 10^{-3}$	68.5
3.58 KB	$1.8 \times 10^{-2}$	65.5
7.16 KB	$3.7 \times 10^{-2}$	62.5

Bits (LSBs). Security analysis showed that our algorithm is secure enough to guarantee secure communication. Our novel algorithm has three levels of security, the huge number of different possible stega-keys, the random position of first pixel where the hiding process starts, and the transposition applied to each 24-bit block of the message to be hidden, all of these make the process of steganalysis more complex even with using a computer works at  $10^6$  extractions/ $\mu s$  which takes  $2.4 \times 10^{24}$  years to exhaustively search a half of the number of stega-keys.

## REFERENCES

- [1] R. Chandramouli and N. Memon, "Analysis of LSB based image steganography," *IEEE ICIP*, pp. 1022–1022, October 2001.
- [2] C. N. M. Pavani, S. Naganjaneyulu, "A survey on LSB based steganography methods," *IJECS*, vol. 2, no. 8, pp. 2464–2467, August 2013.
- [3] M. Juneja and P. S. Sandhu, "An analysis of LSB image steganography techniques in spatial domain," *IJCSEE*, vol. 1, no. 2, 2013.
- [4] W. Stallings, *Cryptography and Network Security - Principles and Practice*, fifth edition ed. Prentice Hal, 2011.
- [5] D. C. Wu and W. H. Tsai., "A steganographic method for images by pixelvalue differencing," *Pattern Recognition Letters*, vol. 24, no. 9-10, p. 16131626, June 2003.
- [6] S. K. L. Gandharba Svalin, "A novel approach to RGB channel based image steganography technique," *International Arab Journal of e-Technology*, vol. 2, no. 4, June 2012.
- [7] M. T. Parvez and A. Gutub, "RGB intensity based variable-bits image steganography," in *Proceedings of 3rd IEEE Asia-Pacific Services Computing Conference (APSCC 2008)*, Yilan, Taiwan, December 9-12 2008.
- [8] A. T. Adnan Gutub, Ayed Al-Qahtani, "Triple-a: Secure RGB image steganography based on randomization," in *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications (AICCSA, 2009)*, Rabat, Morocco, May 10-13 2009, pp. 400–403.
- [9] M. Y. J. Ali Shariq Imran and N. S. Khattak, "A robust method for encrypted data hiding technique based on neighborhood pixels information," *IJCSE*, vol. 1, no. 3, 2007.
- [10] M.-H. S. Jen-Chang Liu, "Generalizations of pixel value differencing steganography for data hiding in images," *Fundamenta Informaticae*, vol. 83, no. 3, pp. 319–335, 2008.
- [11] J. J. Roque and J. M. Minguet, "SLSB: Improving the steganographic algorithm LSB," in *7th International Workshop on Security in Information Systems*, 2009, pp. 57–66.
- [12] A. Ker, "Improved detection of LSB steganography in grayscale images," in *Proc. 6th International Workshop*, vol. 3200. Toronto (Canada): Springer LNCS, May 23-25 2004, p. 97115.
- [13] J. Mielikainen, "LSB matching revisited," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 285–287, 2006.
- [14] D. C. T. Z. . Xiaolong Li, Bin Yang, "A generalization of LSB matching," *IEEE Signal Processing Letters*, vol. 16, no. 2, pp. 69–72, 2009.
- [15] A. Almohammad, "Steganography-based secret and reliable communications: Improving steganographic capacity and imperceptibility," Ph.D., Brunel University, August 2010.

# Survey on Awareness of Privacy Issues in Ubiquitous Environment

Huma Tabassum, Sameena Javaid, Humera Farooq

Department of Computer Sciences  
Bahria University, Karachi Campus  
Karachi, Pakistan

**Abstract**—It is in human nature to keep certain details intimate. This confidential information, if made public, can result in annoyance, humiliation, and even devastating loss. The latter one holds true for individuals as well as organizations. In the case of ubiquitous environment, system adapts to the context of user's actions and intentions and the collection of data becomes pervasive. Thus, privacy turns out to be a vital aspect. This study aims to determine privacy awareness among people in ubiquitous environment. For this purpose, a survey was conducted. This survey was based on a questionnaire. The results show that people consider themselves quite aware, but, more concerned with privacy issues in ubiquitous environment. Also, a significant number of people admit to not taking privacy measures. The analysis was carried out on the basis of current status of participants in the university, namely, Undergraduate students, Graduate students, and Faculty members.

**Keywords**—ubiquitous environment; privacy; privacy awareness

## I. INTRODUCTION

Ubiquitous Computing can be defined as “computing that is omnipresent and is, or appears to be, everywhere all the time; may involve many different computing devices that are embedded in various devices or appliances and operate in the background” [1]. The main focus of this paper is to determine privacy awareness in ubiquitous environment among people. For this purpose, a survey was conducted. The survey was conducted at Bahria University, Karachi campus during October-November, 2014. Participants were asked to fill out a questionnaire. Data samples were collected from students and faculty of various departments, which included Computer Sciences, Computer and Software Engineering, Electrical Engineering, Management Sciences, Humanities, Psychology, and Geophysics departments.

As ubiquitous environment masks the collection of information from the users, this study aims at analyzing awareness about privacy and its consequences. It also investigates the major privacy concerns of users as well as the measures they take. Legalization of privacy issues, improvement in privacy options, and any unpleasant

experience of participants with respect to privacy issues, were also queried. In order to identify potential future dimensions in this area, level of comfort of participants along with the pros and cons of ubiquitous environment were also inquired.

The results indicate that majority of the participants consider themselves quite aware but more concerned with privacy issues; however, a significant number accepts that they do not take countermeasures for privacy issues.

The rest of this paper is organized as follows. The next section describes some of the work on privacy issues in existing research studies. Methodology for carrying out this study is discussed in the section after that. Then the results are discussed, followed by analysis. In the end, conclusion and future work is given which provides some directions to extend this study.

## II. RELATED WORK

Although ubiquitous environment provides ease and continuous access, it makes personal information widely available. Thus, privacy issues emerge as a major concern. Users are unknowingly providing their personal information to the environment which can be misused in case of any malicious attempt [2]. Every technology has both good and bad aspects; this implies that proper measures be taken to ensure privacy of users [15, 16]. It is, therefore, desired that users be aware of not only such systems and devices in their surroundings, but also of the issues these devices pose.

A recent study by Bonné, Quax, and Lamotte [3], shows that most Smartphone users are not truly aware of the privacy threats. Similarly, Renaud, Volkamer, and Renkema-Padmos [4] have also concluded that users are not aware of privacy protection. They tried to analyse mental model of users and made an effort to create understanding of the importance of addressing this issue. In their work, Acquisti, John, and Loewenstein [5] tried to explore how important privacy is to the users. They argued on the rationale of the users and, on how their decisions affect their privacy by analysing the economic model and its effect on privacy issues.

In contrast, a survey based statistical analysis of Hoffman, Novak, and Peralta [6] revealed that approximate 69% of the web users refused to give their data or personal information online to any firm or individual, because they were not certain how their data can be used. Similarly, in another survey based study by Phelps, Novak, and Ferrell [7], 50% respondents claimed transparency about how organizations were using their personal information or individual-specific data. Several theoretical studies also defined that gender makes a great difference regarding privacy concerns while working online or on any social platform [2].

As Angelelli [8] pointed out, major hurdles in providing a uniform method to ensure privacy include the constant advancements in technology, and that the privacy viewpoints differ among people. Thus, there cannot be a fixed number of ways to provide privacy. Usually, privacy issues arise due to improper configuration of a phone and also the network settings [3]. In order to facilitate the users, a recommenders system was proposed by Knijnenburg and Kobsa [9], to allow users to make decisions to disclose information in context.

For many researchers, revelation of the information or data in both aspects (context and use) is the primary fear. The major concerns related to privacy issues include stalking, unnecessary monitoring by employers or government agencies, and location, which can further lead to spam influx [10]. A study by Malhotra, Kim, and Agarwal [11] also presented some core privacy concerns. These were based on data collection and its control regarding privacy.

Privacy regarding personalization has been discussed in several studies [12]. In that study, personalization based on social connections and individual behaviors was analysed. It was shown to assist the programmers in resolving privacy issues. In another study carried out by Anton, Earp, and Young [13], it was revealed that privacy concerns of users have significantly been increasing since 2002.

In order to analyse the stated privacy concerns against the actual behaviour of people, Jiang, Heng, and Choi [14] have argued that people often do not act in accordance with their declared apprehensions. This is consistent with findings of this study that there are differences in what people say and what they actually do. Some new perspectives in online social communications and anonymity to ensure privacy were also discussed in their study [14].

The next section describes the methodology employed in carrying out this study.

### III. METHODOLOGY

In order to achieve the goals of this study, as mentioned in the beginning, a survey was conducted based on a questionnaire. The questions were designed to capture the essence of this study by exploring different approaches to gather the required information from participants. This study also aimed at identifying other potential dimensions related to privacy. These questions may be categorized in following ways.

#### A. General Context Questions

It captured the participants' awareness of ubiquitous environment by some 'feel-good' questions. These questions were helpful in establishing presence of the participants in ubiquitous environment, through careful inquiries of the activities they carry out using their smart device(s) and, their online habits. In addition, some other questions were also included in order to gather basic information, like gender, age group, and current status. These helped to categorize participants for performing analysis.

#### B. Key Privacy Questions

The survey then assessed level of both awareness and concern, of privacy issue in the participants' opinion through questions for each. Some other queries were intended to extract the major privacy concerns and issues of the participants, which came out to be personal information, financial information, and location. Participants were also inquired about the privacy measures they usually take (or not) along with the reasons for not taking any privacy measure. All these questions gave valuable insight to ascertain the behaviours and attitudes of participants with respect to privacy, which was the main objective of this study.

#### C. Trivial Privacy Questions

In order to understand if and why, privacy is a concerning issue, participants were asked about any unpleasant experience they might have encountered. Apart from this, opinion of participants on legalization of privacy issues was also queried. Participants were also inquired about how privacy options should be provided. This aided in establishing premise that such participants should be more conscious about privacy. However, as the results show, this was not always the case.

#### D. Other Dimensions

Finally, to generate further potential dimensions in future, a few questions about level of comfort of participants in ubiquitous environment, as well as the pros and cons of the ubiquitous environment, in general, were also included in the questionnaire. These dimensions can prove to be promising in future, if a thorough study is conducted to analyze the scope of the diverse aspects of privacy concerns and issues.

The next section discusses the results obtained from this study.

## IV. RESULTS

As described in the previous section, this study was based on survey in the form of a questionnaire. These questions were classified in four categories; namely general context questions, key and trivial privacy questions, and other dimensions. On the basis of this analysis criterion, the results obtained are discussed in the following sub-sections. It is important to note that these results reflect on participants' view and have been evaluated and compiled accordingly.

TABLE I  
COMPARISON OF OVERALL RESPONSES WITH EACH CATEGORY

	Overall Response	Categorized Response		
		Undergraduate Students	Graduate Students	Faculty Member
Privacy Awareness	71.4%	68%	74%	63%
Privacy Concern	83.5%	80%	85%	75%

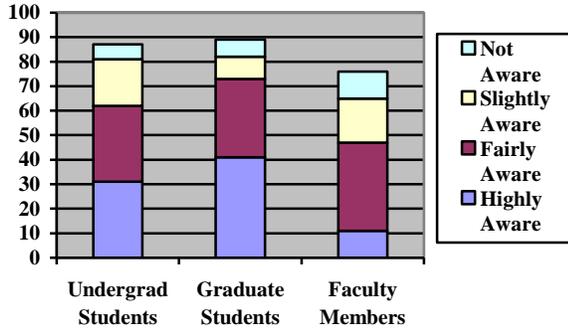


Fig.1 Level of Privacy Awareness in each category of participants

A. Key Privacy Issues

1) *Privacy Awareness:* Among the participants who are part of ubiquitous environment, overall 71.4% considered themselves quite aware about privacy issues, and approximately 10%, not aware. The rest believed their selves somewhat aware of the privacy issues. Categorically, about 68% of the undergraduate students, 74% of graduate students, and 63% of the faculty members consider their selves to be quite aware.

Fig. 1 presents categorical distribution of participants in terms of their position in the university regarding privacy awareness.

2) *Privacy Concern:* After analysing awareness, this study revealed that participants are also concerned of the fact that their data can be used for malicious purpose provided it is easily accessible in ubiquitous environment. All the participants claimed to be concerned about privacy issues to varying extents. A relatively small group of participants (14%) stated to be slightly concerned, while others consider themselves quite concerned.

Fig.2 shows the concern of participants categorically. It should be noted that none of the participants declare that they are un-concerned about privacy.

Among those participants who claim to be concerned about privacy issues, around 80% were undergraduate students, 85% graduate students, and about 75% were faculty members.

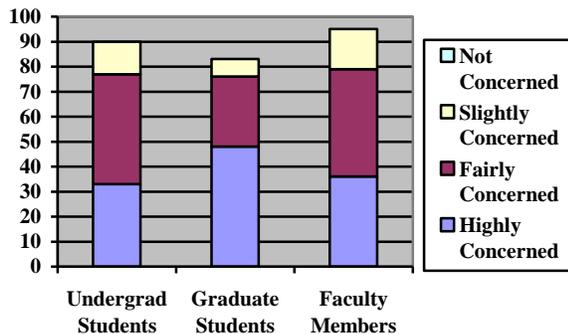


Fig.2 Level of Privacy Concern in each category of participants

Table I encapsulates the responses of participants. Comparison of overall responses against each category is provided. It can be seen from the table that majority of the participants consider themselves quite aware as well as concerned about privacy issues. Another fact emphasized in the table is that the number of participants concerned with privacy issues, are more than those who are aware.

**Pressing Issues:** In this study, it was found that several participants are not only concerned about a single issue; they usually have apprehensions about more than one issue. The major concerns of participants were revealed to be their personal details including friends and family, location, and their financial information respectively.

A very high ratio of 94.5% participants is concerned about their personal details. Among these participants, about 47% were also concerned about their location, and for approximately 41%, their financial information was a concerning issue as well. For about 21% participants, all three of these issues were a privacy concern.

3) *Privacy Measures:* Dealing with privacy issues is another important aspect; however, in this survey, it was found that a considerable fraction of about 19% participants usually do not take any privacy measures. Major reasons for not taking measures are difficulties in configuration as per 65% of the participants. On the other hand, in some of the participants' view (about 4%); there is no need of taking any measures. Among the participants who take countermeasures, majority of about 64% admitted to disabling location with about 37% also modifying application settings with it. Other common privacy measure that approximately 29% participants took was giving false information.

Fig.3 depicts the categorical representation of participants on the basis of whether or not, they take privacy measures.

B. Trivial Privacy Issues

Apart from the key issues analysed above, there were some trivial issues related to privacy that were also explored in this study. These queries provided valuable insight to understand the attitude of participants towards privacy issues.

One such aspect was related to legalization of privacy issues. Most of the participants (about 87%) believe that there should be laws to address privacy issues. However, only about 34% of the participants claimed to be aware of some existing laws.

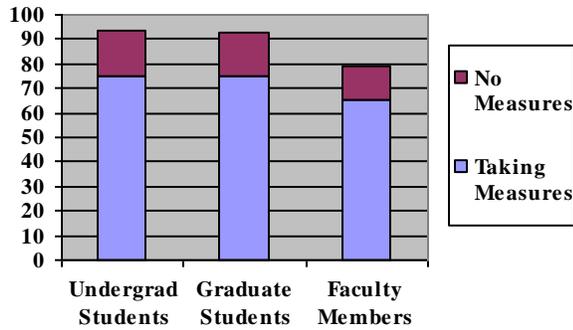


Fig.3 Distribution among participants on the basis taking measures

Among the 36% participants who admitted to have encountered some unpleasant experience due to their non-secure information, a significant 44% claimed to be a victim of fraud.

Another facet, in this study, analysed how the privacy settings can be improved. In the opinion of about 2% of the participants, there is no need for improvement. A very significant group of 57% demanded that more control be given to users over privacy settings.

### C. Other Dimensions

Level of comfort of participants in ubiquitous environment is considered a future dimension in this study. It has been found that approximate 49% participants were quite comfortable with constant monitoring, while about 26% participants felt uncomfortable.

Another aspect, as a potential dimension for future was pros and cons of ubiquitous environment. While taking into account the merits of ubiquitous environment, the chief advantage was ease of life. Another advantage in the view of majority of the participants was fast access, whereas remote monitoring and socialization were also substantial benefits that the participants agreed upon in the survey.

After querying merits of ubiquitous environment, demerits were also questioned in the study. Theft of information and risk of exposure were the major apprehensions of participants, with 74.72% and 61.53% respectively. Other disadvantages in the view of participants were high energy consumption and increased machine reliance.

The next section describes the analysis strategy of this study. It discusses how the results were compiled and how the responses were evaluated.

## V. ANALYSIS AND DISCUSSION

The main goals of this study were to evaluate privacy awareness among people in ubiquitous environment, their level of concern along with major issues, and what privacy measures they take. Apart from these major goals, this survey also investigated some other facets related to privacy as well as ubiquitous environment. For this purpose, data was collected at Bahria University, Karachi campus. The

categorization of participants was done on the basis of their current status in university, that is, Undergraduate students, Graduate students, and Faculty members. The survey was conducted by taking samples of approximately one-fifth of each population in each category.

In order to achieve the above mentioned targets, first thing that needed to be established was whether the participants are part of ubiquitous environment or not. This was done on the basis of general context category of questions. These questions inquired the participants about the use of smart devices, smart appliances, and their usage habits. From the considerable data gathered, about 83% of the participants turned out to be part of ubiquitous environment. In terms of categorization, 86% were undergraduate students, 89% graduate students, and 68% belonged to faculty.

For every research objective, the responses of participants in each category, that is, undergraduate students, graduate students, and faculty, was evaluated separately. The overall percentage of responses has also been provided for comparison. The results were compiled in the light of these considerations.

The next section sums up the findings of this study in a conclusion.

## VI. CONCLUSION

This study has revealed that a fair number of participants (71.4%) claim to be aware of the privacy issues in ubiquitous environment. However, the number of participants who declare their selves concerned about privacy issues, is more than those who claim to be aware (83.5%). Thus, it raises the question that how can people be concerned about something, which they are not aware of in the first place. In addition, despite claiming to be aware and concerned about privacy issues, a considerable number of participants (19%) admit to not taking any countermeasures. This leads to the conclusion that often, people act contrary to what they state.

The next section discusses future work which provides dimensions to extend the study.

## VII. FUTURE WORK

For future work, the data may be analyzed on the basis of gender. This can generate an additional view in the privacy domain. Apart from this, other dimensions briefly mentioned in this study, can be explored further in future. These dimensions include comfort of participants in ubiquitous environment and its pros and cons. These can be used to analyze the scope of the diverse aspects of privacy concerns and issues.

## REFERENCES

- [1] "Glossary of Terms," 1998. [Online]. Available: [http://www.mansfieldct.org/schools/mms/palms/Meet\\_the\\_Team/Glossary.htm](http://www.mansfieldct.org/schools/mms/palms/Meet_the_Team/Glossary.htm). [Accessed 29 November 2014].
- [2] S. Dhaawan, K. Singh and S. Goel, "Impact of Privacy Attitude, Concern and Awareness on use of online social networking," in *5th International Conference-Confluence The Next Generation Information Technology Summit*, 2014.
- [3] B. Bonn e, P. Quax and W. Lamotte, "Your Mobile Phone Is A Traitor! - Raising Awareness on Ubiquitous Privacy Issues with SASQUATCH,"

- International Journal on Information Technologies & Security*, № 3, pp. 39-54, 2014.
- [4] K. Renaud, M. Volkamer and A. Renkema-Padmos, "Why Doesn't Jane Protect Her Privacy?," in *Privacy Enhancing Technologies*, Springer International Publishing, 2014, pp. 244-262
- [5] A. Acquisti, L. John and G. Loewenstein, "What Is Privacy Worth?," *Journal of Legal Studies*, 42(2), pp. 249-274, December 2013.
- [6] D. L. Hoffman, T. P. Novak and M. Peralta, "Building Consumer Trust Online," *Communications of the ACM*, Volume 42 Issue 4, pp. 80-85, 1999.
- [7] J. Phelps, G. Nowak and E. Ferrell, "Privacy concerns and consumer willingness to provide personal information," *Public Policy Marketing*, pp. 27-41, 2000.
- [8] A. Angelelli, *Privacy Issues & Concerns - From a Ubiquitous Point of View*, SIDER, 2007.
- [9] B. P. Knijnenburg and A. Kobsa, "Making Decisions about Privacy: Information Disclosure in Context-Aware Recommender Systems," *ACM Transactions on Interactive Intelligent Systems*, Vol. 3, No. 3, Article 20, 2013.
- [10] L. Barkhuus, B. Brown, M. Bell, S. Sherwood, M. Hall and M. Chalmers, "From awareness to repartee: Sharing Location with social groups," in *Twenty-Sixth annual SIGCHI conference on Human factor in computing systems*, New York, 2008.
- [11] N. K. Malhotra, S. S. Kim and J. Agarwal, "Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model," *Information Systems Research*, pp. 336-355, 2004.
- [12] E. Toch, Y. Wang and L. F. Cranor, "Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems," *Springer's Journal of Personalization Research: User Modeling and User-Adapted Interaction Volume 22, Issue 1-2*, pp. 203-220, 2012.
- [13] A. I. Anton, J. B. Earp and J. D. Young, "How internet users' privacy concerns have evolved since 2002," *Security & Privacy, IEEE*, Volume: 8, Issue: 1, pp. 21-27, 2010.
- [14] Z. Jiang, C. S. Heng and B. C. F. Choi, "Research Note—Privacy Concerns and Privacy-Protective Behavior in Synchronous Online Social Interactions," *Information Systems Research* 24(3), pp. 579-595, 2013.
- [15] M. J. Culnan, "Protecting privacy online: Is self-regulation working?," *Journal of Public Policy & Marketing*, pp. 20-26, 2000.
- [16] R. S. Laufer and M. Wolfe, "Privacy as a concept and a social issue: A multidimensional development theory," *Journal of Social Issues*, Volume 33, Issue 3, pp. 22-42, 1977.

#### AUTHORS PROFILE

HUMA TABASSUM is currently pursuing her Masters degree in Computer Sciences from Bahria University, Karachi. She holds a Bachelors degree in Computer and Information Systems Engineering from N.E.D University of Engineering and Technology, Karachi. Her research interests include data engineering and ubiquitous computing environments.

SAMEENA JAVAID is currently doing her Masters in Computer Sciences from Bahria University, Karachi. She graduated in Computer Sciences from PIMSAT, Karachi. Her current research interests include theoretical foundations and e-learning with context-aware computing.

HUMERA FAROOQ is currently working as an Asst. Professor and Head of Computer Sciences department at Bahria University, Karachi. She did her PhD from Universiti Teknologi Petronas, Malaysia. Her research interests include machine learning, object tracking and detection, and smart environments.

# Efficient Support Vector Machines for Spam Detection: A Survey

Zahra S. Torabi

Faculty of Computer Engineering  
Islamic Azad University of Najafabad  
Isfahan, Iran

Mohammad H. Nadimi-Shahraki

Faculty of Computer Engineering  
Islamic Azad University of Najafabad  
Isfahan, Iran

Akbar Nabiollahi

Faculty of Computer Engineering  
Islamic Azad University of Najafabad  
Isfahan, Iran

**Abstract**— Nowadays, the increase volume of spam has been annoying for the internet users. Spam is commonly defined as unsolicited email messages, and the goal of spam detection is to distinguish between spam and legitimate email messages. Most of the spam can contain viruses, Trojan horses or other harmful software that may lead to failures in computers and networks, consumes network bandwidth and storage space and slows down email servers. In addition it provides a medium for distributing harmful code and/or offensive content and there is not any complete solution for this problem, then the necessity of effective spam filters increase. In the recent years, the usability of machine learning techniques for automatic filtering of spam can be seen. Support Vector Machines (SVM) is a powerful, state-of-the-art algorithm in machine learning that is a good option to classify spam from email. In this article, we consider the evaluation criterions of SVM for spam detection and filtering.

**Keywords**- support vector machines (SVM); spam detection; classification; spam filtering; machine learning;

## I. INTRODUCTION

Influenced by the global network of internet, time and place for communication has decreased by emails. As a result the users prefer to use email in order to communicate with others and send or receive information. In fact spam filtering is an application for classification of emails, and has a high probability of recognizing the spam. Spam is an ongoing issue that has no perfect solution and there is no complete solution technique about spam problem [1]. According to the recent researches done by Kaspersky Laboratory (2014), almost

65.7% of all emails were considered as spam, respectively in January. In this regards, a huge amount of bandwidth is wasted and an overflow occurs while sending the emails. According to reported statistics United State of America, China and South Korea are among the main sources of these spam respectively with 21.9%, 16.0% and 12.5%. Fig. 1 shows the spam sources for each country [2]. Fig. 2 shows the spam sources according to geographical area.

In figure 2, Asia and North America are the greatest sources for the spam, respectively with 49.1 and 22.7 percentage [2]. Recently, separating legitimate emails from spasm has been considerably increased and developed. In fact, separating spam from legitimate emails can be considered as a kind of text classification, because the form of all emails is generally textual and by receiving the spam, the type has to be defined. Support Vector Machines are supervised learning models or out-performed other with associated learning algorithms and good generalization that analyze data and recognize patterns, used for classification and regression analysis. SVM is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. SVM is used to solve quadratic programming problem with inequality constraints and linear equality by discriminating two or more classes with a hyperplane maximizing the margin.

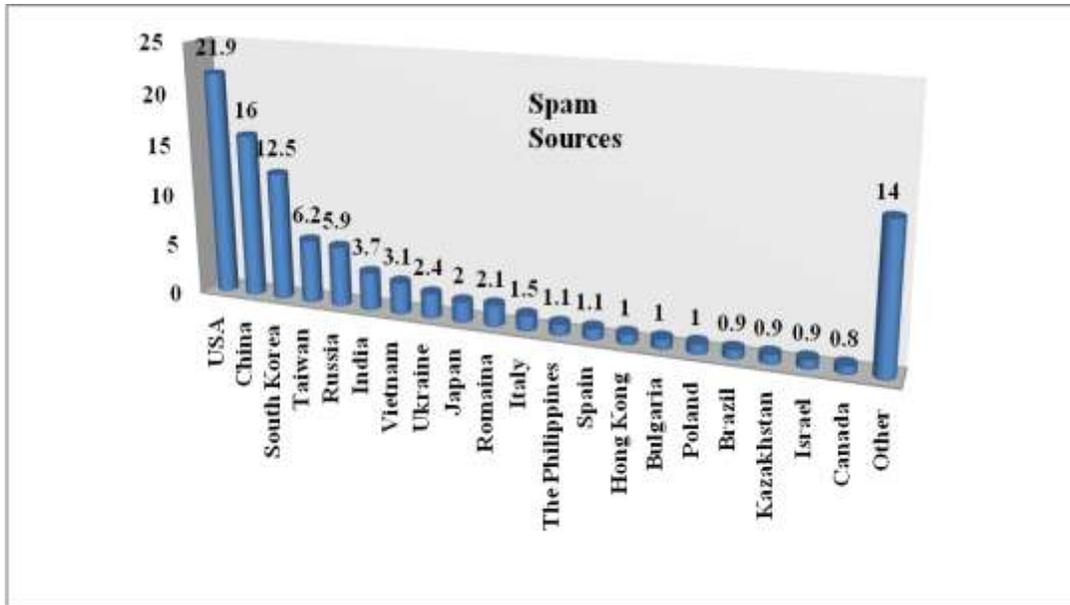


Figure 1. Sources of Spam by country

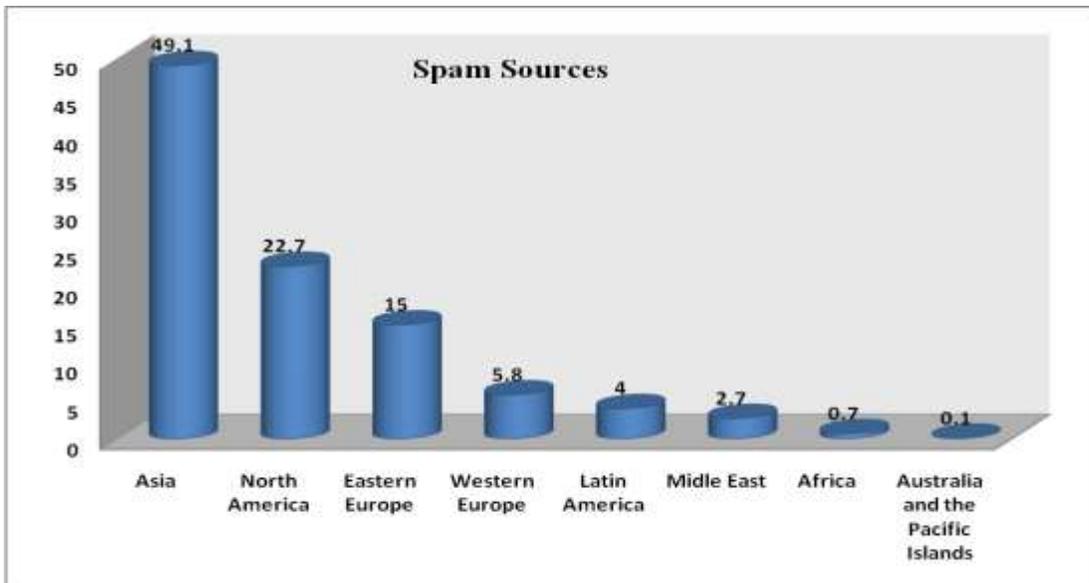


Figure 1. Sources of spam by region

In this paper, we have examined using support vector machine metrics in spam detection. Section 2 discuss an initial background of Spam filtering, discussing spam filtering techniques and content base learning spam filtering architecture. Section 3 introduces standards support vector machine and assessment spam detection using standard support vector machine. Section 4 evaluates spam detection using improved support vector machine. Section 5 is about conclusions and future work.

## II. INITIAL DISCUSSIONS

### A. Spam Filtering Techniques

As the threat is widely spread, variety of techniques have been known to detect spam emails. In fact the techniques are divided two categories Client Side Techniques and Server side techniques [1]. These methods can be applied as a useful filter individually, however in commercial applications, the combination of these methods are generally used in order to recognize the spam more precise. Some of these methods are

defined manually on server side, such as Yahoo email filters. However, they have a main defect and that is the constancy of these pre-defined rules [3]. Another problem is that the spam, spammers can deceive these filters. Furthermore, a popular method of filtering is to identify each spam on the basis of its content [4].

1) *End-User (Client Side Techniques or Techniques to React to Spam)*

These techniques are implemented on client side and once the mails have been downloaded and client examine the mails and then decide what to do with them. So, client can limit the availability of their email addresses, preventing their attractiveness or reducing to spam individually [1, 5-9].

a) *Discretion and Caution*

One way to restrict spam is a shared email address is just among a limited group of co-workers or correspondents. Then, sending and forwarding email messages to receivers who don't know should be rejected.

b) *Whitelists*

A list of contacts that users suppose that they should not send them to the trash folder automatically and they are suitable to receive email from is a whitelist. Then, whitelist methods can also use confirmation or verification [6]. If a whitelist is preemptive, just the email from senders on the whitelist will receive. If a whitelist is not preemptive, it forbids email from being deleted or prevents sent to the spam folder by spam filtering [5]. Usually, just end-users (not email services or Internet service providers) can delete all emails from sources which are not in the white list by setting a spam filter.

c) *Blacklists*

A blacklist method is a mechanism of access control that allow users through except members of the blacklist. It is the opposite method of whitelist [1]. A spam email filtering may save a blacklist of addresses, any message from which would be prevented from achieving its intended destination [5].

d) *Egress spam filtering*

Client can install anti spam filtering to consider emails receiving from the other customers and users as can be done for message coming from the others or rest of the Internet [8].

e) *Disable HTML in email*

Most of programs of email combine web browser or JavaScript practicality, like display of HTML sources, URLs and images. This causes displaying the users to put images in spam easily[9]. On the other hand, there are web bugs on the

spam email written in HTML and it can allow spammers to know the email address is valid.

f) *Port 25 interception*

Network address translation causes rate limiting by intercepting port 25 traffic and direct traffic to mail server and exit spam filtering. On the other hand it can enforce the problems in email privacy, also making it is not able to occur if port 587 submission port isn't used and SMTP-AUTH and STARTTLS is used [7].

g) *Quarantine*

Spam emails are placed in provisional isolation until a proper person like administrator examines them for final characterization [6].

2) *Server side techniques*

In these techniques the server block the spam message. SMTP doesn't detect the source of the message. On the other hand, spammers can forge the source of the message by hijacking their unsecured network servers which is known as "Open relays". Also, "Open proxy", can help the user to forward Internet service request by passing the firewalls, which might block their requests. Verifying the source of the request with open proxy is impossible [1]. Some DNS Blacklists (DNSBLs) have the list of known open relays, known spammers domain names, known proxy servers, compromised zombie spammers, as well as hosts that shouldn't be sending outer emails on the internet, like the address of end-user from a consumer ISP. Spamtraps are usually the email addresses that are invalid or are not valid to collect spam for a long time. A lot of poorly software are written by spammers that cannot right control on the computer sending spam email (zombie computer).Then, they unable to follow with the standards. On the other hand with setting limitations on the MTA<sup>1</sup>, an email server administrator able to decrease spam emails significantly, like enforcing the right fall back of MX<sup>2</sup> records in the Domain Name System, or Teergrube the right controlling of delays' Suffering caused from spam emails is far worse, where some of spam messages might crash the email server temporary.

a) *Limit rate*

This technique restrict the rate of receipt of messages from a user that even this user has not been characterized as a spammer. This technique is used as an indirect way of restricting spam at the ISP level [6].

b) *Spam report feedback loops*

---

<sup>1</sup> mail transfer agent

<sup>2</sup> Mail exchange

ISPs can often prevent seriously damage by monitoring spam reports from places like Spamcop, Network Abuse Clearinghouse and AOL's feedback loop, the domain's abuse@ mailbox, etc to catch spammers and blacklist them [10].

*c) Quantity*

In this technique, in a given time period spam emails detect by examining the number of emails send by a particular user [6, 11]. As the number increases, the possibility that the sender is a spammer increases also.

*d) Domain Keys Identified Mail*

Some systems utilize DNS, as apply DNSBLs to allow acceptance of email from servers which have authenticated in some fashion as senders of only legitimate email but rather than being used to list non conformant sites. [12, 13]. Many authentication systems unable to detect a message is email or spam. Because their lists are static and they allow a site to express trust that an authenticated site will not send spam. Then, a receiver site may select to ignore costly spam filtering methods for emails from the authenticated sites [14].

*e) Challenge/Response*

This technique includes two parties: one party presents a question (“challenge”) and the other party must provide a valid answer or response in order to be authenticated [6]. This method is utilized by specialized services, ISPs and enterprises to detect spam is to need unknown senders by passing different tests before their emails are delivered. The main purpose of this technique is to ensure a human source for the message and to deter any automatically produced mass emails. One special case of this technique is the Turing test and Channel email [15].

*f) Country-based or regional- based filtering*

Some servers of email do not want to communicate with particular regions or countries or from which they receive a great deal of spam. Some countries and regions are mentioned in introduction section according kaspersky lab. Therefore, some email servers use region or country filtering. This technique blocks all emails from particular regions or countries by detecting sender's IP address [16].

*g) Greylisting*

This technique temporarily rejects or blocks messages from unknown sender. In this method for rejecting unknown senders use a 4xx error code that is recognized with all MTAs, so launch to retry delivery later [17]. Greylisting downside is that all legitimate emails from the first time that senders will have a delay time in delivery, with the delay period before a new email is received from an unknown

sender normally being adjustable in the software [18]. It maybe exists that some legitimate emails won't be delivered, that it can happen if a weakly configured but legal mail server realizes the immediate rejection as a stable rejection and sends a bounce email to the main sender, instead of attempting to resend the email later, as it done.

*h) Honey pots*

Another method is really an imitation TCP/IP proxy server which gives the appearance of being an open proxy or simply an imitation mail transfer agent that gives the form of being an open message relay [19]. Spammers who check systems for open proxies/ relays will detect a host and try to send message through it, wasting their resources, time, possibility revealing datas about themselves and also the source of the spam emails sending to the entity that act the honeypot. A system may simply reject the spam efforts, store spams for analysis or submit them to DNSBLs [20].

*i) Sender-supported tags and whitelists*

Some organizations that suggest licensed tags and IP whitelisting which can be placed in message for money to convince receivers' systems that the emails consequently tagged are not as spam email. This system depends on legal implement of the tags. The purpose is for email server administrators to whitelist emails bearing the IP whitelisting and licensed tags [21].

*j) Outbound spam protection*

This method involves detecting spam, scanning email traffic as it exits in network and then taking an action like blocking the email or ignoring the traffic source. Outbound spam protection can be perform on a network-wide level by using policy of routing also it can be run within a standard SMTP router [22]. When the primary economic impact of spam is on sending networks, spam message receivers also experience economical costs, like wasted bandwidth or the risk of having IP addresses rejected by receiving networks. One of the advantage of outbound spam protection is that it blocks spam before it abandons the sending network, maintaining receiving networks globally from the costs and damage. Furthermore it allows system email administrators track down spam email sources in the network such as providing antivirus tools for the customers whose systems have become infected with viruses. Given a suitably designed spam filtering method, outbound spam filtering can be perform with a near zero false positive, that keeps customer related issues with rejected legitimate message down to a minimum [23]. There are some commercial software sellers who suggest specialized outbound spam protection products, such as Commtouch and MailChannels. Open source software like SpamAssassin may be useful.

*k) Tarbits*

A tarpit is a server software that purposely responds very slowly for client commands. With implementing a tarpit that acts acceptable message generally and detect spam email slowly or that seems to be an open mail relay, a site can slow down the rate at which spammers can send messages into the mail simply [24]. Most of systems will really disconnect if the server doesn't answer quickly, which will detect the spam. Then, some legitimate message systems will also do not correctly with these delays [25].

*l) Static content filtering lists*

These techniques require the spam blocking software and/or hardware to scan the entire contents of each email message and find what's inside it. These are very simple but impressive ways to reject spam that includes given words. Its fault is that the rate of false positives can be high, that would forbid someone applying such a filter from receiving legal emails [1, 26]. Content filtering depend on the definition of lists of regular expressions or words rejected in emails. So, if a site receives spam email advertising "free", the administrator may place this word in the configuration of filtering. Then, the email server would reject any emails containing the word [27-29]. Disadvantages of this filtering are divided into 3 folds: Time-consuming is the first one in this filtering. Pruning false positives is Second one. Third one is, the false positives are not equally distributed. Statistical filtering methods use some words in their calculations to know if an email should be classified as email or spam. Some programs that run statistical

filtering are ASSP, DSPAM, Bogofilter, SpamBayes, later revisions of SpamAssassin, Mozilla Thunderbird and Mailwasher.

*m) Content base Learning Spam Filtering Systems*

One of the solutions is an automated email filtering. Many of filtering techniques take usage of machine learning algorithms, that improve their accuracy rate over manual approach. So, many people require filtering intrusive to privacy and also some email administrators prefer rejecting to deny access their machines from sites.[1].Variety of feasible contributions in the case of machine learning have addressed the problem of separating spam emails from legitimate emails [30, 31]. The best classifier is the one that reduces the misclassification rate. However, the researchers have realized later the nature and structure of emails are more than text such as images, links, etc. Some machine learning techniques are such as K-nearest Neighbor classifier, Boosting Trees, Rocchio algorithm, Naïve Bayesian classifier, Ripper and Support Vector Machine [32].

*B. Content base Learning Spam Filtering Architecture*

The common architecture of spam filtering base machine learning or learning content spam filtering is shown in Fig. 3. Firstly, a dataset includes individual user emails which are considered as both spam and legitimate email is needed.

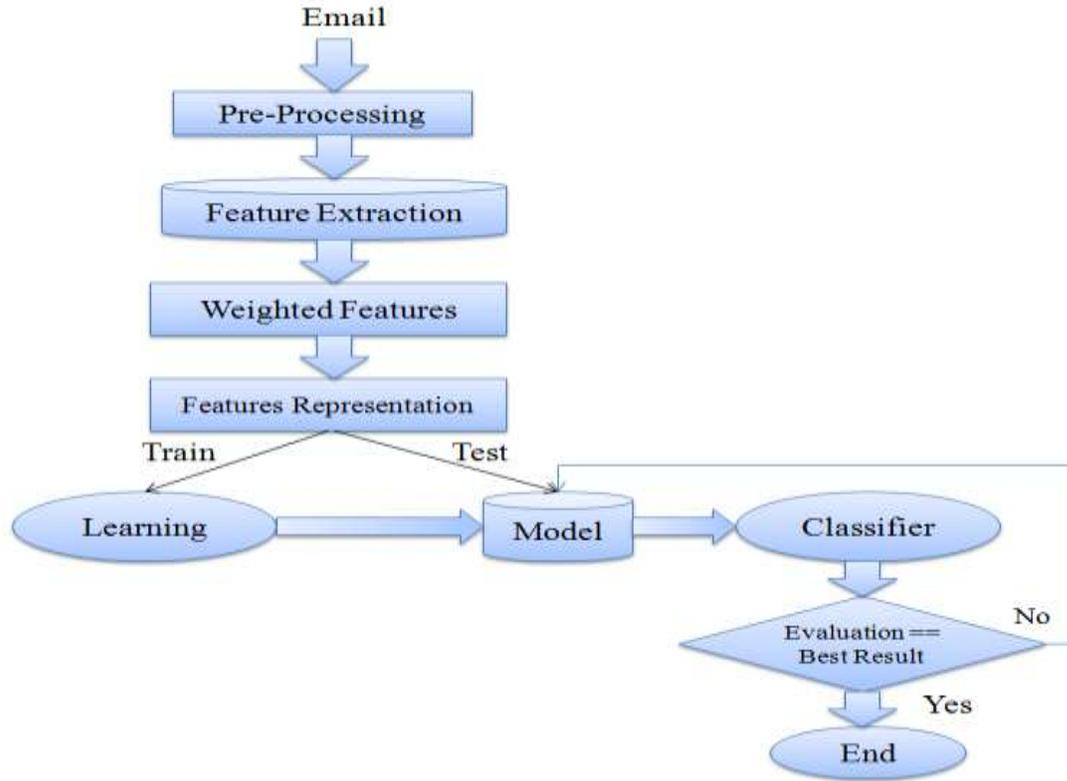


Figure 2. Content base Learning spam filtering architecture

The model includes 6 steps: Pre-Processing, Feature Extraction, Weighted Features, Feature Representation, email data classification, and evaluation or analyzer section. Machine learning algorithms are employed at last to train and test whether the demanded email is spam or legitimate.

#### 1) Pre-processing

When an email receives the first step runs on email is pre-processing. Pre-processing includes tokenization.

##### a) Tokenization

Tokenization is a process to convert a message to its meaningful component. It takes the message and separate it into a series of tokens or words [33]. The tokens are taken from the email's message body, header and subject fields [5]. The tokenization process will extract all the features and words from the message without consideration of its meaning [32].

#### 2) Feature extraction

After pre-processing step and breaks email message into tokens or features. Feature extraction process is started to extract useful features in all features and reduce the space vector. Feature extraction can include stemming, noise removal and stop word removal steps.

##### a) Stemming

Stemming is a method to reduce terms to its main form with stripping the plural from nouns the suffixes from verbs [5, 32]. This process is suggested by Porter on 1980, it determines stemming as an approach for deleting the commoner structural and in-flexional endings from terms in English [34]. A collection of rules is referred to convert words to their stems or roots iteratively. This method increases speed of learning and classification steps for many classes and decrease the number of attribute in the feature space vector [35].

##### b) Noise removal

Unclear terms in an email cause noise. A intentional action of misplaced space, misspelling or embedding particular characters into a word is pointed to as obfuscation. For instance, spammers obfuscated the word Free into "fr33" or Viagra into "V1agra", "V|jagra". Spammers employ this approach in an effort to bypass the right recognition of these words by spam filtering [5, 32] To contrast these misspelled words, Regular expression and statistical deobfuscation techniques is used.

##### c) Stop-word removal

Stop-word removal technique is the process for elimination of usual terms that have the most frequency but have less meaning than the words [36]. Messages contains large number of noninformative words like articles, prepositions, conjunctions and these words will increase the size of the attributes vector space cause complicating the classifier.

### 3) Weighted Features

When the useful features are selected, it is the time to choose a measure to assign the features to create features vectors before classification. Weighted features includes Information gain, Document frequency, Mutual information, Chi-square steps.

#### a) Information gain (IG)

IG is the attribute's impact on reducing entropy [37]. IG calculates the number of bits of information gained for the class by knowing the attendance or lack of presence of a word in a document[5]. Let denote the set of classes in the goal space. IG formula is defined as:

$$G(t) = - \sum_{i=1}^n pr(c_i) \log pr(c_i) \quad (1)$$

$$+ pr(t) \sum_{i=1}^n pr(c_i|t) \cdot \log pr(c_i|t) + pr(\bar{t}) \sum_{i=1}^n pr(c_i|\bar{t}) \cdot \log pr(c_i|\bar{t})$$

#### b) Document frequency [11]

DF points to the number of documents in which an attribute occurs[5].The weight of the attributes is calculated in the lower frequency, that is less than the predefined threshold, is deleted[38]. Negligible attributes that do not contribute to classification are deleted then improving the efficiency of the classifier. The following formula represents the form of document frequency:

$$tf_{ij} = \frac{n_{ij}}{\sum_k n_{kj}} \quad (2)$$

#### c) Mutual information (MI)

MI is a quantity that calculates the mutual dependence of the 2 variables. If a attribute does not depend on a category then it is eliminated from the attribute vector space[39]. For each attribute feature X with the class variable C, MI can be computed as follows:

$$MI(X, C) = \sum_{x,c} P(X=x, C=c) \times \log \frac{P(X=x, C=c)}{P(X=x) \times P(C=c)} \quad (3)$$

MI is an easier method to run and valid in predictions.

#### d) Chi-square

Chi-Square test is a statistical measure that calculates the existence of attribute against the expected number of the existences of those attributes [5]. In the Chi-square test, the features refer to the independent variables and categories indicate the dependent variables that is spam and legitimate [39, 40].

$$X^2(t, c) = \frac{N \times (AD - CB)^2}{(A + C) \times (B + D) \times (A + B) \times (C + D)} \quad (4)$$

Formula 4 calculates the term of goodness, class c and t, which A is the number of times t and c exists together, B is the number of times the exists without c, C is the number of times c exists without t, and D is the number of times neither c not t exists. The chi-square equation for class computation is as follows:

$$X^2_{Avg(t)} = \sum_{i=1}^m pr(c_i) \cdot X^2(t, c_i) \quad (5)$$

$$X^2_{Max(t,c)} = Max_{i=1} \{X^2_{(t,c_i)}\} \quad (6)$$

#### 4) Feature Representation

Feature representation converts the set of weighted features to a specific format required by the machine learning algorithm used. Weighted features are usually shown as bag of words or VSM<sup>1</sup>. The literal features are indicated in either numeric or binary. VSM represents emails as vectors such as  $x = \{x_1, x_2, \dots, x_n\}$ . All features are binary:  $x_i=1$ , if the corresponding attribute is available in the email, differently  $x_i=0$ . The numeric presentation of the features where  $x_i$  is a number represents the occurrence frequency of the attribute in the message. For instance if the word "Viagra" seems in the email then a binary value equal will be assigned to the attribute. The other generally used attribute representation is n-gram character model that gets characters sequences and TF-IDF<sup>2</sup> [41]. N-gram model is N-characters piece of a word. It can be noted as each co-occurring collection of characters in a word. N-gram model infolds bi-gram, tri-gram and qua-gram. TF-IDF is measure that is statistical utilized to calculate how importance a word is to a document in a attribute dataset. Word frequency is defined with TF<sup>3</sup>, that is a number of times the word exists in the email yields the importance of the word in to the

<sup>1</sup> Vector Space Model

<sup>2</sup> frequency-inverse document frequency

<sup>3</sup> term frequency

document. Then frequency is multiplied with  $IDF^1$  which measures the frequency of the word happening in all emails [42].

### 5) Classifier

Supervised machine learning techniques run by collecting the training dataset, which is manually prepared individually. The training dataset have two phases, one phase is legal emails, another one is the spam emails. Then, each email is converted into features that is images, time, words for instance in the email. So, the classifier builds that would determine the nature of next incoming email[1]. Most of machine learning algorithms have been worked in the direction of spam detection like Boosting Trees [43, 44], K-Nearest Neighbor classifier [29, 45], Rocchio algorithm [44, 46], Naïve Bayesian classifier [29, 47-49] and Ripper [44, 50]. Furthermore, these algorithms filter the email by analyzing the header, body or the whole message. Support Vector Machine Is one of the most used techniques as the base classifier to overcome the spam problem [51]. SVM algorithm is used wherever there is a need for model recognition or classification in a specific category or class [52]. Training is somehow easy and in some researches the efficiency is more than other classifications. This is due to the fact that in data training step, Support vectors use data from database but for high dimension data, the validity and efficiency decrees due to the calculating complexities [53, 54].

### 6) Evaluation or Performance Measures

Filtering needs to be evaluated with performance measures that it divided into two categories: decision theory like (false negatives, false positives, true positive and true negative) and information retrieval such as (accuracy, recall, precision, error rate and derived measures) [32]. Accuracy, precision and spam recall are the most practical and useful evaluation parameters. Accuracy indicates the ratio between the number of legitimate mails and the number of correctly classified spam emails to the total correctly classified emails used for testing. Recall refers to the ratio between the number of correctly classified spam against spam that is misclassified as legitimate and the number of spam emails detected as spam. Precision shows the ratio between the numbers of correctly classified spam to the number of all messages recognized as spam. Table 1 represents performance measures of spam filtering:

TABLE I. Performance Measures of Spam Filtering

Performance measures	Equations
<b>Accuracy</b>	$(n_{L \rightarrow L} + n_{S \rightarrow S}) / (n_{L \rightarrow L} + n_{L \rightarrow S} + n_{S \rightarrow L} + n_{S \rightarrow S})$
<b>Error Rate</b>	$(n_{L \rightarrow S} + n_{S \rightarrow L}) / (n_{L \rightarrow L} + n_{L \rightarrow S} + n_{S \rightarrow L} + n_{S \rightarrow S})$
<b>False positive</b>	$(n_{L \rightarrow S}) / (n_{L \rightarrow L} + n_{L \rightarrow S})$
<b>False negative</b>	$(n_{S \rightarrow L}) / (n_{L \rightarrow L} + n_{L \rightarrow S})$
<b>Recall</b>	$(n_{S \rightarrow S}) / (n_{S \rightarrow L} + n_{S \rightarrow S})$
<b>Precision</b>	$(n_{S \rightarrow S}) / (n_{L \rightarrow S} + n_{S \rightarrow S})$
<b>TotalCost Ratio (<math>T_{CR}</math>)</b>	$(n_{S \rightarrow L} + n_{S \rightarrow S}) / \lambda (n_{L \rightarrow S} + n_{S \rightarrow L})$
<b>ROC Curve</b>	Ratio between true positive and false positive for various threshold values.

As shown in table 1,  $n_{L \rightarrow L}$  and  $n_{S \rightarrow S}$  refer to the legitimate emails and spam emails that correctly classified.  $n_{S \rightarrow L}$  indicates spam emails incorrectly have been classified as legitimate emails and  $n_{L \rightarrow S}$  point to the legitimate emails incorrectly have been classified as spam emails. Error rate is the rate between spam and legitimate emails are incorrectly classified to the total correctly classified emails used for testing. False negatives [55] is a measure to recognize the spam emails that are classified as legitimate emails. False positives ( $F_p$ ) refers to the legitimate email classified as. Spam emails that correctly classified as spam emails represent to true positive ( $T_p = 1 - F_N$ ). Legitimate emails that are correctly classified as legitimate emails refer to true negative ( $T_N = 1 - F_p$ ). ROC<sup>2</sup> curve indicates [56], true positive as a function of the false positive for various threshold values. Total cost ratio ( $T_{CR}$ ) use for comparing the rate of effectual filtering by a given value of  $\lambda$  in comparison with no filtering is used. If  $T_{CR} > 1$ , the using of filtering is proper.

## III. EVALUATION OF SPAM DETECTION WITH STANDARD SVM

### A. Standard Support Vector Machines

SVM is a classifier that is included as a sub branch of Kernel Methods in Machine Learning. It is based on statistical learning theory [57]. In SVM, assuming a linear categories are separate frames of acceptance, with a maximum margin hyperplane that acquires are separate categories. Separate data linearly in matters that are not accepted frames of data are mapped to a higher dimensional space so that they can be separated in the new space linearly. If there are two categories that are linearly separable, what is the best way of separating of two

<sup>1</sup> inverse document frequency

<sup>2</sup> Received Operating Characteristic

categories? Various algorithms such as "Perceptron" can do this separation. The idea of SVM, to separate a category is to create two borders: two parallel plate boundary classification and draw them away from each other so that to hit the data we are dealing. Board Categories that owns the maximum distance from the boundary of the plates may be best the separator. Fig. 4 shows SVM classification:

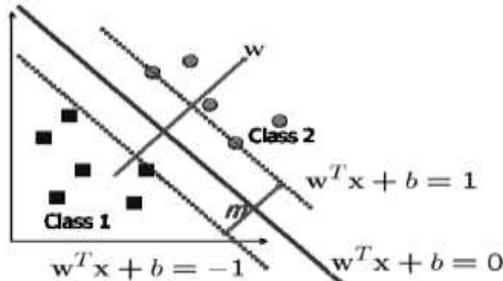


Figure 3. SVM classification

Nearest training data to the hyperplane separator plates is called "support vector." The proper using of the SVM algorithm is the power of generalization, because despite having large dimensions, the over fitting can be avoided. This property is caused by the optimization algorithm used in data compression. In fact, Instead of training data, support vectors are used. To solve the problem, we have a number of training samples,  $x \in \mathfrak{R}^n$  which is a member of the class  $y$  and  $y \in [-1, 1]$ . SVM linear decision function is defined as follows:

$$f(x) = \text{sign}(\langle w, x \rangle + b), w \in \mathfrak{R}^n, b \in \mathfrak{R} \quad (7)$$

Separating hyperplane is defined below:

$$\langle w, x \rangle + b = 0, w_1 x_1 + w_2 x_2 \dots + w_n x_n + b = 0 \quad (8)$$

We should find the values of  $W, b$  so that the training samples are closely grouped, with the assumption that data can be separated linearly to a maximum margin. For this purpose we use the following equation:

$$\text{Max: } W(\alpha) = -\frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j x + \sum_{i=1}^N \alpha_i \quad (9)$$

$$\text{subject to: } \sum_{i=1}^N \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, \dots, N \quad (10)$$

In the relations of (9) and (10), the values of  $\alpha$  or Dual Coefficient and  $b$  by using QP equations are solved. New values of  $x$ , the test phase in relation to the following places:

$$\mathcal{F}(x, \alpha, b) = wx + b = \sum_{i=1}^N \alpha_i y_i x + b = \sum_{i=1, \dots, n} \alpha_i y_i x + b \quad (11)$$

There are different kernel functions. Thus, solving equations using kernel function is defined as follows:

$$\mathbf{n - th Degree Polynomial: } K(x, x') = (1 + \langle x, x' \rangle) \quad (12)$$

$$\text{Neural Network: } K(x, x') = \tan(k_1 \langle x, x' \rangle + k_2) \quad (13)$$

$$\text{Radial basis: } K(x, x') = \exp(-\|x - x'\|^2 / c) \quad (14)$$

SVM can be used in pattern recognition and where you need to identify patterns or categories of classes particular. Training is fairly simple. Compromise between complexity and classification error rate clearly is controlled. Fig. 5 presents SVM algorithm [58].

---

**ALGORITHM 1 : Support Vector Machine**

---

**Input:** sample  $x$  to classify, Training set  
 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ;  
**Output:** decision  $y_j \in [-1, 1]$   
compute SVM solution  $w, b$  for data set with imputed labels  
compute outputs  $y_i = (w \cdot x_i) + b$  for all  $x_i$  in positive bags  
set  $y_i = \text{sign}(y_i)$  for every  $i \in [-1, 1]$   
**for** (every positive bag  $B_i$ )  
**if** ( $\sum_{i \in I} (1 + y_i)/2 = 0$ )  
  Compute  $i^* = \arg \max_{i \in I} y_i$   
  Set  $y_i = 1$   
**end**  
**end**  
**While** (imputed labels have changed)  
  Output ( $w, b$ )  
  Put values  $w, b$  in  $y_i$  and get the result  $y_i$

---

Figure 4. SVM algorithm

### B. Examining of Standard SVM in Spam Detection

A spam email classification based on SVM classifier was presented by Drucker and his co-workers. In this article, speed and accuracy metrics on SVM and 3 other classifications like Ripper, Rocchio and Boosting Decision trees were compared on 2 datasets. One dataset had 1000 features another one had over than 7000 features. Moreover, TF-IDF and applying STOP words were used on features. Results had been shown that the speed of binary SVM algorithm on training dataset was much higher than other classification and its accuracy was very near to Boosting algorithm. Also the error rate of binary SVM was lower than other classification on two dataset with 0.0213% [44]. Woitaszek et al used linear SVM classification and a personalized dictionary to model the training data. They proposed a classification that implemented on Microsoft Outlook XP and they could categorize the emails on Outlook XP. The accuracy rate of proposed method was 96.69% that was 1.43 higher than system dictionary with the accuracy rate of 95.26 [59]. Matsumoto and et al applied their tests results on two classifications such as SVM and Naive Bayesian and used TF and TF-IDF on features vector. The accuracy rate of 2 classifiers were the same, but the one with a lower false alarm rate and miss rate is better classifier. Their results showed that Naive Bayesian has better performance than SVM on a dataset. The false alarm rates and miss rates for Naive Bayesian were stable in almost all the data sets [60]. Scheffer et al, developed an approach for learning a classifier using publicly available

(labeled) and (unlabeled) messages. In this case  $n$  users were currently subscribed, the classifier for a new user was obtained by training a linear SVM, where the misclassification cost of the messages was based on the estimated bias-correction term of each message. The experiments run on Enron corpus and Spam messages from various public and private sources of email users on binary representation. It was verified that the proposed formulation ( $1 - \text{AUC}$ ) decreased the risk up to 40%, in comparison with the user of a single classifier for all user [61]. Kanaris et al, used  $n$ -gram characters while  $n$  was predefined and was used as a variable in linear SVM. In this research, information retrieval was used to select features and the binary representation words frequency TF were also used. Experiment run on LingSpam and Spam Assassin data sets. 3 - , 4 - and 5 - gram and 10-fold cross-validation was performed.  $N$ -gram model is better than the other methods. Results were shown that with variable  $N$  in a cost-sensitive scenarios, with binary features, spam precision seem to provide better results. Also TF features are better than for spam recall. Then, spam precision rate was higher than 98% and spam recall rate was higher than 97%. TCR value of the proposed approach was not greater than 1 because the precision ratio failed to be 100% [62]. Ye et al provided a distinct model based on SVM and DS theory. They used SVM to classify and sort mail based on header contents features and applied DS theory to detect spammers with the accuracy rate of 98.35% [63]. Yu and Xu et al, compared 4 machine learning algorithms such as NB<sup>1</sup>,

---

<sup>1</sup> Naive Bayes

NN<sup>1</sup>, SVM and RVM<sup>2</sup>. Test results showed that NN classifier was very sensitive and was the only unfit for rejecting spam mails. SVM, RVM had much better performance than NB. RVM had a higher run time rather than SVM [53]. Chhabra et al used SVM classifier for spam filtering and compared multiple VM kernel function on the Enron data set with different rates. In these tests, the performance of the linear kernel and polynomial kernel with degree= 1 were equal because the performance of those kernel functions were the same. If the degree of polynomial kernel increases, the performance of its function decreases [64]. Shahi et al used SVM and Naïve Bayes to classify the Nepali SMS as non-Spam and Spam. The accuracy measure was used to empirical analyze for various text cases to evaluate of the classification methodologies. The accuracy of SVM was 87.15% and the accuracy of Naïve Bayes was 92.74% [65]. A method based on Artificial Immune System (AIS) and incremental SVM were proposed for spam detection. In this study to effect change of the content of email dynamically and label email a sliding window was used. Experiments run on PU1 and Ling datasets with compared 8 methods, considered as different including Hamming Distance(without and with without mutation), SVM ,Included Angle and Weighted Voting. The results show that methods in SVM group had a good performance with miss rate below 5% on corpus PU1. When the number of support vectors was growth, the speed would increase. The performance of AIS group has been unsatisfactory. On the other hand with the increase in size of window from 3 to 5, the performance of WV group and AIS group raised [66]. Summary of this section is shown in Table 2.

TABLE II. OVERVIEW OF STANDARD SVM IN SPAM EMAILS

Idea	year	Authors
In this study, 3 classification such as Ripper, Rocchio and Boosting Decision trees were compared with SVM. Moreover, TF-ID was used for both binary features, Bag of Words. STOP words was applied on features. Results represented that the performance in term of speed and accuracy of binary SVM was much higher than other classification and its accuracy was very close to Boosting algorithm.	1999	Harris Drucker, , Donghui Wu, and Vladimir N. Vapnik
To model the training data, used linear SVM classification and a linear dictionary. The classification implemented on Microsoft Outlook XP and they could categorize the emails on Outlook. The accuracy rate of proposed method was 96.69%.	2003	Woitaszek M, Shaaban M, Czernikowski R

In this research SVM and Naive Baysiean compared together. TF and TF-IDF were applied on features. Their results showed that Naive Baysian had better performance than SVM on a dataset.	2004	Matsumoto R, Zhang D, Lu M
In this research, n-gram characters, information retrieval were used to select features. Binary representation words frequency TF were also applied on features. Experiments run on LingSpam and Spam Assassin data sets with 3 -, 4 - and 5 - gram and 10-fold cross-validation. N-gram model is better than the other methods and results show that spam precision rate was higher than 98% and spam recall rate was higher than 97%. TCR value of the proposed approach was not greater than 1.	2007	Kanaris, I., Kanaris, K., Houvardas, I., & Stamatatos, E
A model is presented a classification framework for learning generic messages, available (labeled) and unavailable messages. Linear SVM was used for classification a new user. Experimental results implemented on Enron corpus dataset with binary representation. This research proposed the risk formula (1 - AUC) decreased to 40%.	2007	Bickel, S, Scheffer, T
A distinct model based on SVM and DS theory was suggested. They used SVM to classify and sort mail based on header contents features and applied DS theory to detect spammers with the accuracy rate of 98.35%.	2008	Ye M, Jiang QX, Mai FJ
4 machine learning algorithms NB, NN, SVM and RVM compared. Test results showed that NN was very sensitive and unfit for rejecting spam mails. SVM, RVM had much better performance than NB and RVM had a higher run time rather than SVM.	2008	Yu B, Xu Z
SVM classifier used for spam filtering and compared multiple VM kernel function on the Enron data set with different rates. In these tests, the performance of the linear kernel, polynomial kernel with degree one were equal. If the degree of the polynomial kernel increases, performance decreases.	2010	Priyanka Chhabra, Rajesh Wadhvani, Sanyam Shukla
Naïve Bayes and SVM were used to classify the Nepali SMS as Spam and non-Spam and it was found to be 87.15% accurate in SVM and 92.74% accurate in the case of Naïve Bayes.	2013	Tej Bahadur Shahi, Abhimanu Yadav
A hybrid method based on AIS and incremental SVM were suggested. A sliding window to effect the change of the content of email dynamically and label email was used. Experiments run on PU1 and Ling datasets with compared 8 methods. The results show that SVM group had a good performance with miss rate below 5% on corpus PU1. When the number of support vectors was growth, the speed would increase. The performance of AIS group has been unsatisfactory. On the other hand with the increase in size of window from 3 to 5, the performance of WV group and AIS group had raised.	2014	TAN, Y and RUAN, G

<sup>1</sup> neural network

<sup>2</sup> Relevance vector machine

#### IV. EVALUATION OF SPAM DETECTION WITH IMPROVED SVM

##### A. Improved Support Vector Machines

One common weakness from the point of parametric methods like SVM classification is that the computational complexity is not appropriate for data with high dimensions. Weight ratio is not constant, so the marginal value is varied. Also there is need to decide to choose a good kernel function in order to select and the proper value for C parameter. SVM is suitable to deal with the problems of limited training data features [67]. In researches of [55, 68-72], SVM is much more efficient than other non-parametric classifications for example Neural Networks, K-Nearest Neighbor in term of classification's accuracy, computational time and set the parameters but it has a weak function in the data set with high dimension features. Four classifications including neural networks, SVM, J48, simple Bayesian filtering were used for spam email data set. All emails were classified as spam (1) or not (0). Compared with the J48 and simple Bayesian classifier with many features, it was reported that SVM, neural network and NN did not show a good result. Based on this fact, the researchers concluded that the NN and SVM are not suitable for classification of large datasets email. The result of the study of [64] revealed that SVM captures a high range of time and memory for big size of the data set. To solve the classification problem of SVM, the most effective and proper features are necessary as feature candidates rather than using the entire feature space and the sample choose as support vectors in order to keep performance and accuracy for SVM.

##### B. Examining of Improved SVM in Spam Detection

Wang et al proposed a new hybrid algorithm based on SVM and Genetic algorithms to select the best email features named GA-SVM and then compared it with SVM on UCI spam database. The experiments were represented that the new algorithm is more accurate. The accuracy rate for proposed method was 94.43 that 0.05 increased rather than SVM with accuracy rate 94.38% [73]. Ben Medlock & Associates introduced a new adaptive method named "ILM<sup>1</sup>" which used the combination of Weights and N-gram language and in the end compared it with SVM, BLR<sup>2</sup>, MNB<sup>3</sup>. The results showed that the ILM accuracy is higher than other algorithms with 0.9123%

and SVM accuracy was 0.8472% [74]. A new approach based Online SVM was proposed for spam filtering which was compatible with any new training set, for each system. In this method, an adaptive situation was provided for the parameter C (one of the main issues in the classification of SVM, the choice to obtain the maximum distance between a bunch of C) and then it was compared with SVM method. The proposed method accuracy was 0.1% more than SVM accuracy [75]. Blanco et al suggested a solution to reduce false negative errors based on SVM for spam filtering. In this study a combining of multiple dissimilarities of an ensemble of SVM is proposed. Results had shown that the proposed method was more efficient rather than SVM [76]. Blanzieri et al, had an improvement on the SVM classifier to detect spam by localizing data. In this reasearch, two algorithms were proposed and implement on TREC 2005 Spam Track corpus and SpamAssassin corpus, one was the SVM Nearest Neighbor Classifier which was a combination of SVM and K Nearest Neighbor and the second one was HP-SVM-NN which was the previous algorithm with a high degree of probability. Both methods were compared with SVM and the results show that the accuracy of these two algorithms are higher than SVM with 0.01% higher[77]. Sun et al used two algorithms LPP<sup>4</sup> and LS-SVM<sup>5</sup> in order to detect spam. They used LPP algorithm for extracting features from emails and LS-SVM algorithm for classification and detection of spams from mails received. Their results showed that the performance of proposed method was better than the other categories with the accuracy rate of 94% [78]. Tseng et al proposed an Incremental SVM for spam detection on dynamic social networks. The proposed system was called MailNET that installed on the network. Several features extracted from user for training on dataset of the network were applied and then updating plan with the incremental learning SVM. The proposed system implemented on a data set from a scale of university email server. Results have shown that the proposed MailNET was effective and efficient in real world [79]. Ren proposed email spam filtering framework for feature selection using SVM classifier. In this way, (TF-IDF) weight was applied on features. The accuracy of proposed method on TREC05p-1, TREC06p and TREC07p datasets were 98.830, 99.6414% and 99.6327%. Experiments represent that the proposed method of feature extraction increases the effectiveness and performance of text detection with less computational cost and the proposed model can run on dataset in other languages, such as Japanese ,Chinese etc [80]. Rakse et al used SVM classifier for spam filtering and proposed a new kernel function called Cauchy kernel

---

<sup>1</sup> Interpolated Language Model

<sup>2</sup> Bayesian Logistic Regression

<sup>3</sup> Multinomial Naive Bayes

---

<sup>4</sup> locality pursuit projection

<sup>5</sup> least square version of SVM

function. Experiments run on ECML-PKDD dataset and results represented that the performance of the new kernel function was better AUC values on experiments of eval01,eval02 and eval03 datasets with the accuracy of 0.72343%, 0.77703%, 0.89118% when  $C=1.0$  [81]. Yuguo et al prepared a sequential kernel functions commonly called PDWSK to classify SVM. The kernel function had the ability to identify dependence criteria among existing knowledge when words created on the net and capable to calculate the semantic similarity in a text and had higher accuracy compared with SVM. The proposed method was run on trec07p corpus with 5-cross validation and compared with other kernel function of SVM such as RBF, Polynomial, SSK, WSK. The precision, recall and F1 measures for PDWSK were 93.64%, 92.21%, 92.92% that were higher than the other kernel functions [82]. A predictive algorithm combined with fuzzy logic, Genetic algorithms and SVM classifier with RBF kernel was presented. The proposed method used LIBSVM and MATLAB to implemented SVM, fuzzy rules and GA. The proposed method can detect errors in pages according to their SVM classification in comparison with standard SVM. The accuracy of proposed method had a higher efficiency with 95.6% [83]. Hsu and Yu proposed a combination algorithm from Staelin and Taguchi methods for the optimization of SVM with the choice of parameters for classifying spam email. The proposed method SVM ( $L64$  ( $32 \times 32 \times 2$ )) were compared with other methods such as improved Grid search(GS), SVM(Linear), Naïve Bayes and SVM(Taguchi Method  $L32$ ) on 6 data sets of Enronspam Corpa. If the parameters  $C$  and  $\gamma$  were not set up for linear kernel SVM, the accuracy of SVM(Linear) would lower than proposed method and Naïve Bayes algorithms. On the other hand, the proposed method was not the best accuracy and was lower than  $GS(32 \times 32)$ . But GS need computing  $32 \times 32 = 1024$  times, searching and the proposed method required 64 times. The propose method was 15 times faster. The accuracy of proposed method was near to GS and can select good parameters for SVM with RBF kernel [84]. FENG and ZHOU proposed two algorithms OCFS<sup>1</sup>, MRMR<sup>2</sup> for dimension reduction and elimination related features. They incorporated OCFS and MRMR and proposed OMFS<sup>3</sup> algorithm. This algorithm had two phases: the first phase was OCFS algorithm which selected features from data space and used in the next stage. The second phase of the algorithm used the characteristics of the candidates features which are selected then MRMR was used to reduce the redundant attributes. These algorithms reduce the dimensions on the

classifications of Naive Bayes, KNN and SVM on PUI dataset. Results have shown that MRMR get the best feature selection accuracy in comparison of SVM, NB and KNN. The worst accuracy rate belongs to CHI. The accuracy of NB on CHI is lower than 85%. It also displays the poorer accuracy on SVM, with fluctuating on 75%. The accuracy of KNN has increased with feature selection but even below 85%. Finally, with increasing feature of the proposed algorithm the accuracy, F-Measure, ROCA on proposed method increases in comparison of the other algorithms [54]. Maldonado and Huillier, proposed a distributed feature selection method on datasets with minimal error determined nonlinear decision boundaries. Also by using two-dimensional formulation, can reduce the number of features are useless on SVM binary. With proposed method, the width of RBF kernel is optimized by using of the reduced gradient. Experiments run on 2 real world spam datasets. Results show that the proposed feature selection method perform better than the other feature selection algorithms when a smaller number of variables are using [85]. Yang et al used LSSVM<sup>4</sup> algorithm to detect spam. These algorithms solved the problem of garbage tags. In this method, inconsistent changes in the structure of traditional SVM, is converted to a balanced structure. An empirical function for square errors exists in the test data set. So Quadratic Program (QP), convergent to the linear equations. This algorithm increases the speed and classification accuracy under high dimensional gisette\_scale data set. LSSVM training time was near less 10 times than SVM. The accuracy of SVM was 47.50% and the LSSVM accuracy was 60.50% [86]. Hong-liang ZHOU and LUO proposed a method by combining SVM and OCFS<sup>5</sup> for feature selection algorithm to detect spams. Experiments were performed on 5 spam corpuses (ZH1, PU1, PU2, PU3 and PUA). The result showed that the proposed method compared with other traditional combinations, had better performance in terms of accuracy and F-Measure. The accuracy rate of proposed method was above 90% on 5 dataset of spam corpuses [87]. GAO et al modify the SVM classifier by exploiting web link structure. They firstly construct the link structure preserving within-class scatter matrix with direct link matrix and indirect link matrix. Then they incorporate web link structure into SVM classifier to reformulate an optimization problem. The proposed method is useful for the link information on the web. Results show that the combination of web link structure and SVM can significantly outperform related methods on web spam datasets. The proposed method almost had better performs from the other method of

---

<sup>1</sup> Orthogonal Centroid Feature Selection

<sup>2</sup> Minimum Redundancy Maximum Relevance

<sup>3</sup> Orthogonal Minimum Feature Selection

---

<sup>4</sup> Least Squares support vector machine classifiers

<sup>5</sup> Orthogonal Centroid Feature Selection

SVM on WEBSpAM-UK2006 except spam pages accuracy, followed by MCLPVSVM and MCVSVM. Also, the results on link features were better than results on other feature selection combination. It was clear that the proposed method had better performs on WEBSpAM-UK2007 integral features. Although, the proposed method was just a little better than MCVSVM, it maybe had the reason that the indirect link matrix and the direct link matrix were both sparse matrices [88]. Renuka and et.al proposed a method name Latent Semantic Indexing (LSI) for feature extraction method to select the proper and suitable feature space. The Ling spam email corpus datasets was used for the experimentation. The accuracy of SVM (TF-IDF) was 85% while the accuracy of SVM (LSI) was 93% Then the performance improvement of SVM (LSI) over the SVM (TF-IDF) was 8% [89] The Summary of this section is shown in table 3.

TABLE III. OVERVIEW OF IMPROVED SVM IN SPAM EMAILS

Authors	Year	Idea
Huai-bin Wang, Ying Yu, and Zhen Liu	2005	A new hybrid algorithm (GA-SVM) is proposed based on SVM and Genetic algorithm. GA used to select suitable email features. The proposed method was compared with SVM. Results show that the new algorithm was more accurate with 0.05 increased.
Ben Medlock, William Gates Building, JJ Thomson Avenue	2006	A new adaptive method named ILM was proposed which used the combination of Weights and n-gram language. ILM compared with SVM, BLR and MNB. The results showed that the ILM accuracy is higher than other algorithms with 0.9123%.
D.Sculley, Gabriel M. Wachman	2007	A new approach based Online SVM was proposed for spam filtering which was compatible with any new training set, for each system. In this method, an adaptive situation was provided for the parameter C. The proposed method had better performance rather than SVM and its accuracy was 0.1% more than SVM.
Angela Blanco, Alba María Ricket, Manuel Martín-Merino	2007	A solution to reduce false negative errors based on SVM for spam filtering is suggested. Then to achieve this goal an ensemble of SVM that hybrids multiple dissimilarities is proposed. Results have shown that the proposed method is more efficient rather than SVM with one branch.
Enrico Blanzieri, Anton Bryl	2008	Two algorithms were proposed, one was the SVM Nearest Neighbor Classifier which was a combination of SVM and K Nearest Neighbor and the second one was HP-SVM-NN which is the previous algorithm with a high degree of probability. Results show that the accuracy

		of these two algorithms were higher than SVM with 0.01% higher.
Sun X, Zhang Q, Wang Z	2009	Two algorithms LP Pand LS-SVM were proposed. LPP algorithm used for feature selection and LS-SVM algorithm used for classification. The results showed that the performance was better than the other categories with the accuracy rate of 94%.
Chi-Yao Tseng, Ming-Syan Chen	2009	An Incremental SVM for spam detection on dynamic social networks named MailNET was suggested. The proposed system was installed on the network. Several features extracted from user for the training of the network were applied and then updating plan for the incremental learning SVM.
Qinqing Ren	2010	An email spam filtering framework for feature selection using SVM classifier was proposed. In this way, the attribute frequency (TF-IDF) weight is applied on features. The accuracy of proposed method on TREC05p-1, TREC06p and TREC07p datasets were 98.830, 99.6414% and 99.6327% and proposed model can run on datasets in other languages, such as Japanese ,Chinese etc.
Surendra Kumar Rakse,Sanyam Shukla	2010	A new kernel function for SVM classifier in spam detection was proposed, called Cauchy kernel function and the performance measured on ECML-PKDD dataset. Results are shown that the performance of the new kernel function is better than rest.
Liu Yuguo , Zhu Zhenfang, Zhao Jing	2011	A sequential kernel functions commonly to classify SVM called DPWSK was proposed. DPWSK kernel can identify dependence criteria among existing knowledge and can calculate the semantic similarity in a text and had higher accuracy compared with SVM. Results show that precision=93.64%, recall=92.21% and F1=92.92% for PDWSK.
S. Chitra, K.S. Jayanthan, S. Preetha, R.N. Uma Shankar	2012	A predictive hybrid algorithm with fuzzy logic, GA and SVM classifier was presented. The proposed algorithm can detect errors in pages according to fuzzy rules and GA and can classifier with SVM classification. Results show the accuracy of SVM had a higher efficiency with 95.6%.
Wei-Chih Hsu , Tsan-Ying Yu	2012	A combination Algorithm with Staelin and Taguchi methods to aim optimization of SVM and the choice of parameters for classifying spam email, have been proposed. The parameters of the proposed method with other methods such as improved grid search on 6 data sets were compared. The results show that the propose method was 15 times faster than GS and the accuracy of proposed method was near to GS.

Yu FENG, Hongliang Zhou	2013	A hybrid algorithms base on OCFS and MRMR for dimension reduction named OMFS was proposed. OMFS had 2 phases: first OCFS algorithm run to select features from data space second MRMR to reduce the redundant attributes. These algorithms reduce the dimensions of Naive Bayes, KNN and SVM on PU1 dataset. Results have shown that with increasing feature of the proposed algorithm, the accuracy, F-Measure and ROCA on these classification have been increased.
Sebastián Maldonado, Gaston L’Huillier	2013	A distributed approach on datasets with minimal error was determined nonlinear decision boundaries. Also used two-dimensional formulation to reduce the number of features on SVM binary. With proposed method, the width of RBF kernel is optimized by using of the reduced gradient. Results on 2 real spam dataset represented that the proposed feature selection method perform better than the other feature selection algorithms when a smaller number of variables were used.
Xiaolei Yang , Yidan Su , JinPing Mo	2013	LSSVM algorithm is proposed to solve the problem of garbage tags. In this method, quadratic program (QP), convergent to the linear equations with inconsistent changes in the structure of traditional SVM that is converted to a balanced structure. Also an empirical function for square errors exists in the test data set. This idea increases speed and classification accuracy. LSSVM training time was near less10 times than SVM. The accuracy of SVM was 47.50% and the LSSVM accuracy was 60.50%
Hong-liang Zhou, Changyong Luo	2014	A hybrid method base on and OCFS for feature selection is proposed. Experiment results are run on five spam corpuses (PU1, PU2, PU3, PUA and ZH1). The result showed that F-Measure and accuracy of proposed method are more excellent than other traditional combinations. The accuracy rate of proposed method was above 90% on 5 dataset of spam
Shuang Gao, Huaxiang Zhang, Xiyuan Zheng, Xiaonan Fang	2014	A framework to modify the SVM classifier by exploiting web link structure is proposed. They firstly construct the link structure preserving within-class scatter matrix with direct link matrix and indirect link matrix. Then they incorporate web link structure into SVM classifier to reformulate an optimization problem.
Renuka, K.D. , Visalakshi, P	2014	A method name Latent Semantic Indexing (LSI) for feature extraction is proposed. The Ling spam email corpus datasets was used for the experimentation. The accuracy of SVM (TF-IDF) was 85% while the accuracy of SVM (LSI) was 93%.

## V. CONCLUSIONS AND FURTHER WORK

When spam email come to internet, they become a problem for Internet users to present a conservative estimate of 70 to 75 percent of email-related products. The most dynamic and best methods of machine learning techniques in spam filtering, is a high-speed filtering with high accuracy. In this paper we review and examine support vector machine to detect and classifier spam as standard and improved with combined with other classification algorithms, dimension reduction and improved with different kernel functions. SVM algorithm is suitable for pattern recognition, classification, or anywhere that needs to be classified in a special class, can be used. In some studies, its performance relative to other categories more thrust, because the data in the data training phase of support vectors are selected .In the computational complexity of high-dimensional data collection, the performance decrease, so it can be classified by algorithms reduce the size and selection of features to be combined or select good value for it's parameters like C and  $\gamma$  that some of them are mentioned in this article.

## REFERENCES

- [1] Amayri, O., *On email spam filtering using support vector machine*. 2009, Concordia University.
- [2] kaspersky. 2014; Available from: <http://www.kaspersky.com/about/news/spam/>.
- [3] Cook, D., et al. *Catching spam before it arrives: domain specific dynamic blacklists*. in *Proceedings of the 2006 Australasian workshops on Grid computing and e-research-Volume 54*. 2006. Australian Computer Society, Inc.
- [4] Zitar, R.A. and A. Hamdan, *Genetic optimized artificial immune system in spam detection: a review and a model*. Artificial Intelligence Review, 2013. **40**(3): p. 305-377.
- [5] Subramaniam, T., H.A. Jalab, and A.Y. Taqa, *Overview of textual anti-spam filtering techniques*. International Journal of Physical Sciences, 2010. **5**(12): p. 1869-1882.
- [6] Nakulas, A., et al. *A review of techniques to counter spam and spit*. in *Proceedings of the European Computing Conference*. 2009. Springer.
- [7] Seitzer, L., *Shutting Down The Highway To Internet Hell*. 2005.
- [8] Du, P. and A. Nakao. *DDoS defense deployment with network egress and ingress filtering*. in *Communications (ICC), 2010 IEEE International Conference on*. 2010. IEEE.
- [9] Sheehan, K.B., *E - mail survey response rates: A review*. Journal of Computer - Mediated Communication, 2001. **6**(2): p. 0-0.
- [10] Rounthwaite, R.L., et al., *Feedback loop for spam prevention*. 2007, Google Patents.
- [11] Sandford, P., J. Sandford, and D. Parish. *Analysis of smtp connection characteristics for detecting spam relays*. in

- Computing in the Global Information Technology, 2006. ICCGI'06. International Multi-Conference on. 2006. IEEE.
- [12] Allman, E., et al., *DomainKeys identified mail (DKIM) signatures*. 2007, RFC 4871, May.
- [13] Delany, M., *Domain-based email authentication using public keys advertised in the DNS (DomainKeys)*. 2007.
- [14] Leiba, B. and J. Fenton. *DomainKeys Identified Mail (DKIM): Using Digital Signatures for Domain Verification*. in CEAS. 2007.
- [15] Iwanaga, M., T. Tabata, and K. Sakurai, *Evaluation of anti-spam method combining bayesian filtering and strong challenge and response*. Proceedings of CNIS, 2003. 3.
- [16] Dwyer, P. and Z. Duan. *MDMap: Assisting Users in Identifying Phishing Emails*. in *Proceedings of 7th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS)*. 2010.
- [17] Heron, S., *Technologies for spam detection*. Network Security, 2009. 2009(1): p. 11-15.
- [18] González-Talaván, G., *A simple, configurable SMTP anti-spam filter: Greylists*. Computers & Security, 2006. 25(3): p. 229-236.
- [19] Spitzner, L., *Honeypots: tracking hackers*. Vol. 1. 2003: Addison-Wesley Reading.
- [20] Dagon, D., et al. *Honeystat: Local worm detection using honeypots*. in *Recent Advances in Intrusion Detection*. 2004. Springer.
- [21] Ihalagedara, D. and U. Ratnayake, *Recent Developments in Bayesian Approach in Filtering Junk E-mail*. SRI LANKA ASSOCIATION FOR ARTIFICIAL INTELLIGENCE, 2006.
- [22] Goodman, J., G.V. Cormack, and D. Heckerman, *Spam and the ongoing battle for the inbox*. Communications of the ACM, 2007. 50(2): p. 24-33.
- [23] Goodman, J.T. and R. Rounthwaite. *Stopping outgoing spam*. in *Proceedings of the 5th ACM conference on Electronic commerce*. 2004. ACM.
- [24] Hunter, T., P. Terry, and A. Judge. *Distributed Tarptitting: Impeding Spam Across Multiple Servers*. in *LISA*. 2003.
- [25] Agrawal, B., N. Kumar, and M. Molle. *Controlling spam emails at the routers*. in *Communications, 2005. ICC 2005. 2005 IEEE International Conference on*. 2005. IEEE.
- [26] Zdziarski, J.A., *Ending spam: Bayesian content filtering and the art of statistical language classification*. 2005: No Starch Press.
- [27] Khorsi, A., *An overview of content-based spam filtering techniques*. Informatica (Slovenia), 2007. 31(3): p. 269-277.
- [28] Obied, A., *Bayesian Spam Filtering*. Department of Computer Science University of Calgary amaobied@ucalgary.ca, 2007.
- [29] Androutsopoulos, I., et al., *Learning to filter spam e-mail: A comparison of a naive bayesian and a memory-based approach*. arXiv preprint cs/0009009, 2000.
- [30] Seewald, A.K., *An evaluation of naive Bayes variants in content-based learning for spam filtering*. Intelligent Data Analysis, 2007. 11(5): p. 497-524.
- [31] Sebastiani, F., *Machine learning in automated text categorization*. ACM computing surveys (CSUR), 2002. 34(1): p. 1-47.
- [32] Guzella, T.S. and W.M. Caminhas, *A review of machine learning approaches to spam filtering*. Expert Systems with Applications, 2009. 36(7): p. 10206-10222.
- [33] Zdziarski, J., *Tokenization: The building blocks of spam*. Ending Spam: Bayesian Content Filtering and the Art of Statistical Language Classification, 2005.
- [34] Porter, M.F., *An algorithm for suffix stripping*. Program: electronic library and information systems, 1980. 14(3): p. 130-137.
- [35] Ahmed, S. and F. Mithun. *Word Stemming to Enhance Spam Filtering*. in CEAS. 2004. Citeseer.
- [36] Silva, C. and B. Ribeiro. *The importance of stop word removal on recall values in text categorization*. in *Neural Networks, 2003. Proceedings of the International Joint Conference on*. 2003. IEEE.
- [37] Kent, J.T., *Information gain and a general measure of correlation*. Biometrika, 1983. 70(1): p. 163-173.
- [38] Tokunaga, T. and I. Makoto. *Text categorization based on weighted inverse document frequency*. in *Special Interest Groups and Information Process Society of Japan (SIG-IPJS)*. 1994. Citeseer.
- [39] Yang, Y. and J.O. Pedersen. *A comparative study on feature selection in text categorization*. in *ICML*. 1997.
- [40] Yezaurunis, W.S., et al. *A unified model of spam filtration*. in *Proceedings of the MIT Spam Conference, Cambridge, MA, USA*. 2005.
- [41] Ramos, J. *Using tf-idf to determine word relevance in document queries*. in *Proceedings of the First Instructional Conference on Machine Learning*. 2003.
- [42] Church, K. and W. Gale, *Inverse document frequency (idf): A measure of deviations from poisson*, in *Natural language processing using very large corpora*. 1999, Springer. p. 283-295.
- [43] Carreras, X. and L. Marquez, *Boosting trees for anti-spam email filtering*. arXiv preprint cs/0109015, 2001.
- [44] Drucker, H., S. Wu, and V.N. Vapnik, *Support vector machines for spam categorization*. Neural Networks, IEEE Transactions on, 1999. 10(5): p. 1048-1054.
- [45] Blanzieri, E. and A. Bryl. *Instance-Based Spam Filtering Using SVM Nearest Neighbor Classifier*. in *FLAIRS Conference*. 2007.
- [46] Rocchio, J.J., *Relevance feedback in information retrieval*. 1971.
- [47] Androutsopoulos, I., et al., *An evaluation of naive bayesian anti-spam filtering*. arXiv preprint cs/0006013, 2000.
- [48] Androutsopoulos, I., et al. *An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with personal e-mail messages*. in *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*. 2000. ACM.

- [49] Metsis, V., I. Androutsopoulos, and G. Paliouras. *Spam filtering with naive bayes-which naive bayes?* in CEAS. 2006.
- [50] Cohen, W.W. *Learning rules that classify e-mail*. in AAAI Spring Symposium on Machine Learning in Information Access. 1996. California.
- [51] Schölkopf, B. and A.J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. 2002: MIT press.
- [52] Schölkopf, B. and A.J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond (adaptive computation and machine learning)*. 2001.
- [53] Yu, B. and Z.-b. Xu, *A comparative study for content-based dynamic spam classification using four machine learning algorithms*. Knowledge-Based Systems, 2008. **21**(4): p. 355-362.
- [54] FENG, Y. and H. ZHOU, *An Effective and Efficient Two-stage Dimensionality Reduction Algorithm for Content-based Spam Filtering\**. Journal of Computational Information Systems, 2013. **9**(4): p. 1407-1420.
- [55] Chapelle, O., P. Haffner, and V.N. Vapnik, *Support vector machines for histogram-based image classification*. Neural Networks, IEEE Transactions on, 1999. **10**(5): p. 1055-1064.
- [56] Fawcett, T., *ROC graphs: Notes and practical considerations for researchers*. Machine learning, 2004. **31**: p. 1-38.
- [57] Vapnik, V.N. and V. Vapnik, *Statistical learning theory*. Vol. 2. 1998: Wiley New York.
- [58] Andrews, S., I. Tsochantaridis, and T. Hofmann. *Support vector machines for multiple-instance learning*. in *Advances in neural information processing systems*. 2002.
- [59] Woitaszek, M., M. Shaaban, and R. Czernikowski. *Identifying junk electronic mail in Microsoft outlook with a support vector machine*. in *2012 IEEE/IPSJ 12th International Symposium on Applications and the Internet*. 2003. IEEE Computer Society.
- [60] Matsumoto, R., D. Zhang, and M. Lu. *Some empirical results on two spam detection methods*. in *Information Reuse and Integration, 2004. IRI 2004. Proceedings of the 2004 IEEE International Conference on*. 2004. IEEE.
- [61] Bickel, S. and T. Scheffer, *Dirichlet-enhanced spam filtering based on biased samples*. Advances in neural information processing systems, 2007. **19**: p. 161.
- [62] Kanaris, I., et al., *Words versus character n-grams for anti-spam filtering*. International Journal on Artificial Intelligence Tools, 2007. **16**(06): p. 1047-1067.
- [63] Ye, M., Q.-X. Jiang, and F.-J. Mai. *The Spam Filtering Technology Based on SVM and DS Theory*. in *Knowledge Discovery and Data Mining, 2008. WKDD 2008. First International Workshop on*. 2008. IEEE.
- [64] Chhabra, P., R. Wadhvani, and S. Shukla, *Spam filtering using support vector machine*. Special Issue IJCCT, 2010. **1**(2): p. 3.
- [65] Shahi, T.B. and A. Yadav, *Mobile SMS Spam Filtering for Nepali Text Using Naïve Bayesian and Support Vector Machine*. International Journal of Intelligence Science, 2013. **4**: p. 24.
- [66] Tan, Y. and G. Ruan, *Uninterrupted approaches for spam detection based on SVM and AIS*. International Journal of Computational Intelligence, 2014. **1**(1): p. 1-26.
- [67] Auria, L. and R.A. Moro, *Support vector machines (SVM) as a technique for solvency analysis*. 2008, Discussion papers//German Institute for Economic Research.
- [68] Kim, K.I., et al., *Support vector machines for texture classification*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002. **24**(11): p. 1542-1550.
- [69] Wei, L., et al., *A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications*. Medical Imaging, IEEE Transactions on, 2005. **24**(3): p. 371-380.
- [70] Song, Q., W. Hu, and W. Xie, *Robust support vector machine with bullet hole image classification*. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 2002. **32**(4): p. 440-448.
- [71] Kim, K.I., K. Jung, and J.H. Kim, *Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2003. **25**(12): p. 1631-1639.
- [72] Youn, S. and D. McLeod, *A comparative study for email classification*, in *Advances and Innovations in Systems, Computing Sciences and Software Engineering*. 2007, Springer. p. 387-391.
- [73] Wang, H.-b., Y. Yu, and Z. Liu, *SVM classifier incorporating feature selection using GA for spam detection*, in *Embedded and Ubiquitous Computing-EUC 2005*. 2005, Springer. p. 1147-1154.
- [74] Medlock, B. *An Adaptive, Semi-Structured Language Model Approach to Spam Filtering on a New Corpus*. in CEAS. 2006.
- [75] Sculley, D. and G.M. Wachman. *Relaxed online SVMs for spam filtering*. in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. 2007. ACM.
- [76] Blanco, Á., A.M. Rickett, and M. Martín-Merino, *Combining SVM classifiers for email anti-spam filtering*, in *Computational and Ambient Intelligence*. 2007, Springer. p. 903-910.
- [77] Blanzieri, E. and A. Bryl, *E-Mail Spam Filtering with Local SVM Classifiers*. 2008.
- [78] Sun, X., Q. Zhang, and Z. Wang. *Using LPP and LS-SVM for spam filtering*. in *Computing, Communication, Control, and Management, 2009. CCCM 2009. ISECS International Colloquium on*. 2009. IEEE.
- [79] Tseng, C.-Y. and M.-S. Chen. *Incremental SVM model for spam detection on dynamic email social networks*. in *Computational Science and Engineering, 2009. CSE'09. International Conference on*. 2009. IEEE.
- [80] Ren, Q. *Feature-fusion framework for spam filtering based on svm*. in *Proceedings of the 7th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference*. 2010.

- [81] Rakse, S.K. and S. Shukla, *Spam classification using new kernel function in support vector machine*. 2010.
- [82] Yuguo, L., Z. Zhenfang, and Z. Jing, *A word sequence kernels used in spam-filtering*". Scientific Research and Essays, 2011. **6**(6): p. 1275-1280.
- [83] Chitra, S., et al., *Predicate based Algorithm for Malicious Web Page Detection using Genetic Fuzzy Systems and Support Vector Machine*. International Journal of Computer Applications, 2012. **40**(10): p. 13-19.
- [84] Hsu, W.-C. and T.-Y. Yu, *Support vector machines parameter selection based on combined Taguchi method and Staelin method for e-mail spam filtering*. International Journal of Engineering and Technology Innovation, 2012. **2**(2): p. 113-125.
- [85] Maldonado, S. and G. L'Huillier, *SVM-Based Feature Selection and Classification for Email Filtering*, in *Pattern Recognition-Applications and Methods*. 2013, Springer. p. 135-148.
- [86] Yang, X.L., Y.D. Su, and J.P. Mo, *LSSVM-based social spam detection model*. Advanced Materials Research, 2013. **765**: p. 1281-1286.
- [87] Hong-liang ZHOU and C.-y. LUO, *Combining SVM with Orthogonal Centroid Feature Selection for Spam Filtering*, in *International Conference on Computer, Network*. 2014. p. 759.
- [88] GAO, S., et al., *Improving SVM Classifiers with Link Structure for Web Spam Detection\**. Journal of Computational Information Systems, 2014. **10**(6): p. 2435-2443.
- [89] Renuka, K.D. and P. Visalakshi, *Latent Semantic Indexing Based SVM Model for Email Spam Classification*. Journal of Scientific & Industrial Research, 2014. **73**(7): p. 437-442.

# A Proposed Approach for Monitoring Quality of Web Services Using Service Level Agreement

Nagy Ramadan Darwish

Department of Computer and Information Sciences, Institute of Statistical Studies and Research, Cairo University, Cairo, Egypt

Rabab Emad Mohamed

Department of Computer and Information Sciences, Institute of Statistical Studies and Research, Cairo University, Cairo, Egypt

Doaa Hany Elsayed

Department of Computer and Information Sciences, Institute of Statistical Studies and Research, Cairo University, Cairo, Egypt

**Abstract**— Web service technology has gained more important role in developing distributed applications and systems on the Internet. Rapid growth of published web services makes their discovery more and more difficult. Nowadays, most of web service providers sign Services Level Agreement (SLA) contracts with their clients in order to guarantee the offered functionality of their services. This paper proposes an approach to monitor Quality of Services (QoS) in web service according to Service Level Objectives (SLO) in SLA. Monitoring procedures are introduced to check variations in the pre-agreed metric values of SLAs. Then, the deviation between the actual quality and the acceptable quality level can be identified and analyzed. Finally, the weaknesses of the web service practices can be discovered and solved.

**Keywords:** Web Services, Services Level Agreement, Quality of Services, Web Service Level Agreement, Services Level Agreement Metric.

## I. INTRODUCTION

A web service is a software system designed to support interoperable machine-to-machine interaction over a network [1,2, 3]. Other systems interact with the web service using Simple Object Access Protocol (SOAP) messages, typically conveyed using HTTP with an XML serialization in conjunction with other web-related standards [3,4, 5]. The web services architecture is based upon the interactions between three primary roles: service provider, service registry, and service requestor [6]. These roles interact using publish, find, and bind operations [7]. The service provider is the business that provides access to the web service and publishes the service description in a service registry [7]. The service requestor finds the service description in a service registry and uses the information in the description to bind to a service [7, 8]. Service registry is a searchable registry of service descriptions where service providers publish their service descriptions [7]. A logical view of the web services architecture is shown in Figure 1.

Web services are composed of functional and non-functional attributes. The non-functional attributes are referred to as QoS. QoS is defined in [9], adapting from the definition of quality in ISO 8402, as a set of non-functional attributes of the entities used in the path from a web service repository to the consumer who relies on the ability of a web service to satisfy its stated or implied needs in an end-to-end fashion. Some examples of QoS attributes are performance, reliability, security, availability, usability, discoverability, and adaptability [10, 11].

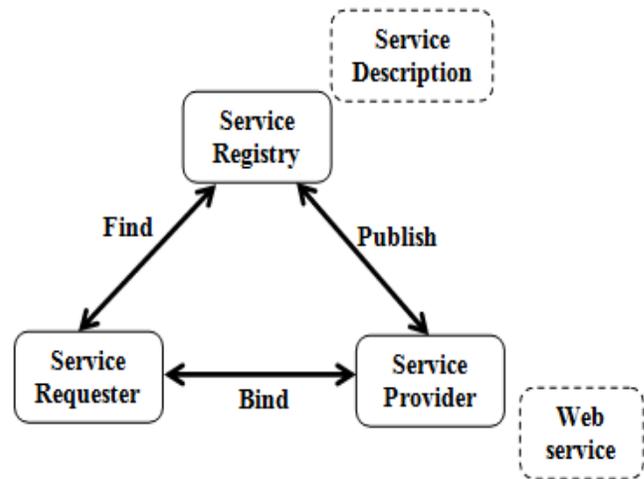


Figure (1): Web Services Architecture [8]

A QoS attribute of web service may have a metric for quantification that can be used to assess the service's performance [12]. This metric is specified in the web service contract between service partners. Measure provides a quantitative indication of the extent, amount, dimensions, capacity, or size of some attributes of a product or process [13]. Monitoring is the process of observing quality of the service over a period of time and to find the degree of deviation from the expected norm [14]. In the case of

monitoring statically composed services, corrective actions for SLA violations can be incorporated only by stopping the execution and resuming it with an alternate concrete composition [14].

SLA is a contract between a network service provider and a customer that specifies what services the network service provider will provide. Many Internet Service Providers (ISP)s provide their customers with an SLA [8]. SLA faces many challenges such as: a lack of accurate input, involvement and commitment from the business and customers, the tools and resources required to agree, document, monitor, report and review agreements and service levels, the process becomes a bureaucratic, administrative process rather than an active and proactive process delivering measurable benefit to the business, business and customer measurements are too difficult to measure and improve, so are not recorded, inappropriate business and customer contacts and relationships are developed, high customer expectations and low perception, poor and inappropriate communication is achieved with the business and customers [15].

This paper proposes a proposed approach for monitoring QoS of SLA in web services. This approach used event-based monitoring and history-based monitoring. It includes many activities within SLA and checking actual metrics with pre-defined to detect variations.

The rest of this paper is divided into 6 sections. Section (II) explores Related Work. Section (III) defines SLA definition, importance, life cycle and Web Service Level Agreement (WSLA). Section (IV) discusses QoS Model. Section (V) introduces the proposed monitoring approach. Section (VI) summarizes conclusion and future work.

## II. RELATED WORK

Many researchers in this domain focus on main points: monitoring of web services, SLA management including QoS management and mapping techniques of monitored metrics to SLA parameters and attributes.

- Karthikeyan. J et al, propose a frame work called Quality based Dynamic Composition (QDC) framework. This framework includes ontology's for improving the web semantic discovery and also provides the best service according to the customer specification as well as monitoring the execution of web service and evaluating the QoS parameters associated with web service [8].
- Katawut Kaewbanjong et al, review attributes, categorization, and metric of QoS of web services. In addition, create a new classification scheme where each category under this scheme can point to the information needed for successful application of QoS attributes [13].
- Shanmuga Priya R et al, propose an approach to monitor QoS properties of compositions that are dynamically composed. Aspect Oriented Programming (AOP) based

approach is adapted for monitoring and triggering alternate service composition, in case of SLA violations. The proposed approach is demonstrated for an application of foreign exchange market that involves composing services at runtime [14].

- G. Dobson et al, present a unified QoS ontology applicable to the main scenarios identified such as QoS-based web services selection, QoS monitoring and QoS adaptation [16].
- S. Ran, proposes a QoS model and a Universal Description, Discovery and Integration (UDDI) extension for associated QoS to a specific web service. The author does not specify how QOS Model values are actually assessed and monitored. It is assumed that they are specified by the service provider in UDDI [17].
- M. Comuzzi, proposes architecture for monitoring of SLAs considering two requirements introduced by SLA establishment: the availability of historical data for evaluating SLA offers and the assessment of the capability to monitor the terms in an SLA offer [18].
- H. G. Song et al, propose a simulation based web service performance analysis tool that calls the web service once under low load conditions and then transform these testing results into a simulation model [19].
- D. Gunter et al, present a NetLogger that is a distributed monitoring system, which monitors and collects information from networks. Applications can invoke NetLogger's API to survey the overload before and after some request or operation. However, it monitors only network resources [20].
- T. Suzumura et al, work on performance optimizations of web services. Their approach is to minimize XML processing time by using differential deserialization [21].
- B. Koller et al, discuss autonomous QoS management using a proxy-like approach. SLAs can be exploited to define certain QoS parameters that a service has to maintain during its interaction with a specific customer. However, their approach is limited to web services and does not consider requirements of cloud computing infrastructures like scalability [22].
- M. Tian et al, present an approach for integrating QoS with web services. The authors implemented a tool-suite for associating, querying and monitoring QoS of a web service [23].
- I. Brandic et al, present an approach for adaptive generation of SLA templates. SLA users can define mappings from their local SLA templates to the remote templates in order to facilitate communication with numerous cloud service providers. However, they do not investigate mapping of monitored metrics to agreed SLAs [24].

- I. Brandic et al, deal with QoS attributes for web services. They identified important QoS attributes and their composition from resource metrics. They presented some mapping techniques for composing QoS attributes from resource metrics to form SLA parameters for a specific domain. However, they did not deal with monitoring of resource metrics [25].
- F. Rosenberg et al and L. Zeng et al, propose a QoS model and a middleware approach for dynamic QoS-driven service composition. They investigate a global planning approach to determine optimal service execution plans for composite service based on QoS criteria [26,27].
- L. Zeng et al, introduce a model-driven approach for integrating performance prediction into service composition processes carried out using BPEL. They composed service SLA parameters from resource metrics using some mapping techniques. But they did neither consider resource metrics – nor SLA monitoring [28].

### III. SERVICE LEVEL AGREEMENT (SLA)

SLA sets the expectations between the consumer and provider [29]. It helps to define the relationship between the two parties. It manages how the service provider sets and maintains commitments to the service consumer [6]. A properly specified SLA describes each service offered and addresses:

- How delivery of the service at the specified level of quality will become realized
- Which metrics will be collected
- Who will collect the metrics and how
- Actions to be taken when the service is not delivered at the specified level of quality and who is responsible for doing them
- Penalties for failure to deliver the service at the specified level of quality
- How and whether the SLA will evolve as technology changes (e.g., multi-core processors improve the provider's ability to reduce end-to-end latency) [30]

SLAs can be either static or dynamic. A static SLA is an SLA that generally remains unchanged for multiple service time intervals. A dynamic SLA is an SLA that generally changes from service period to service period, to accommodate changes in provision of service. Any SLA may contain the following parts: purpose, parties, validity period, scope, restrictions, service-level objectives (availability, performance, and reliability), optional services, administration authority [31]. In the following sub-sections, the researchers explain SLA importance, SLA life cycle, and how to formulate SLAs in the web services.

#### A. SLA Importance:

SLA is important [32] because it sets boundaries for the following aspects of service provisioning.

- Customer commitments: Focused on customer requirements and assure that the internal processes follow the right direction.
- Key performance indicators for the customer service: Improved customer satisfaction stays a clear objective.
- Key performance indicators for the internal organizations: Internal objectives become clearer and easier to measure.
- The price of non-conformance: If the SLA has penalties non-performance can be costly. However, by having penalties defined, the customer understands that the provider truly believes in its ability to achieve the set performance levels.

#### B. SLA Life cycle

SLAs have a certain life cycle that consists of six phases [33] as shown in figure 2.

1. Service and SLA Template Development: This phase includes the identification of (service consumer needs, appropriate service characteristics and parameters) that can be offered given the service execution environment, and the preparation of standard SLA templates.

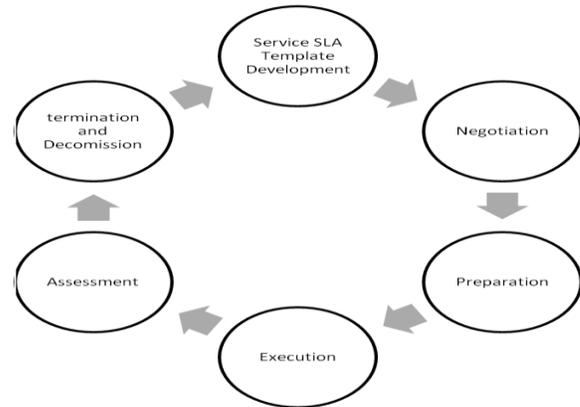


Figure (2): SLA Life Cycle

2. Negotiation: This phase includes the negotiation of the specific values for the defined service parameters, the costs for the service consumer, the costs for the service provider when the SLA is violated, and the definition and periodicity of reports to be provided to the service consumer.
3. Preparation: The service is prepared for consumption by the service consumer. This phase may require the reconfiguration of the resources that support service execution in order to meet SLA parameters.
4. Execution: This phase is the actual operation of the service. It includes service execution and monitoring,

real-time reporting, service quality validation, and real-time SLA violation processing.

5. **Assessment:** This phase has two parts:
  - Assessment of the SLA and the QoS that is provided to an individual consumer. QoS, consumer satisfaction, potential improvements, and changing requirements are reviewed periodically for each SLA.
  - Assessment of the overall service. This assessment can be tied to an internal business review. Elements to be covered in this review are the QoS provided to all consumers, the need for the realignment of service goals and operations, the identification of service support problems, and the identification of the need for different service levels.
6. **Termination and Decommission:** This phase deals with termination of the service for reasons such as contract expiration or violation of contract. Also, this phase deals with decommission of discontinued services.

### C. Expressing SLAs in the Web Service Level Agreement Language

WSLA [34] was developed by IBM and is used to define SLA documents. A WSLA is an agreement between a service provider and a customer and as such defines the obligations of the parties involved. It is the obligation of a service provider to perform a service according to agreed-upon guarantees for the service parameters on the technical level. The design goals of WSLA are a formal and flexible XML-based language for SLA definitions between different organizations. WSLA is structured in three sections: the Parties section, the Service Description and Obligations section. The Parties section [35] identifies the contractual parties and contains the technical properties of a party, i.e. their address and interface definitions. The service description section [35] specifies the characteristics of the service and its observable parameters as follows: For every service operation, one or more binding. In addition, one or more SLA Parameters of the service may be specified. Obligations Section [35] defines the SLOs, guarantees and constraints that may be imposed on the SLA parameters. This allows the parties to unambiguously define the respective guarantees they give each other. The WSLA language provides two types of obligation: Service Level Objectives represent promises with respect to the state of SLA parameters and Action Guarantees are promises of a signatory party to perform an action. This may include notifications of service level objective violations or invocation of management operations.

## IV. SLA QUALITIES

There are two categories of qualities that can be specified in SLAs [31, 33]: measurable and unmeasurable. Figure 3 shows the two categories of SLA qualities.

### A. Measurable Qualities

Measurable qualities can be measured automatically using metrics; for example, the percentage of time a system is available. Measurable qualities include accuracy, availability, capacity, cost, latency, provisioning-related time, reliable messaging, and scalability.

- **Accuracy:** is concerned with the error rate of the service. It is possible to specify the average number of errors over a given time period.
- **Availability:** is concerned with the mean time to failure for services. It is possible to specify
  - The system's response when a failure occurs
  - The time it takes to recognize a malfunction
  - How long it takes to recover from a failure
  - Whether error handling is used to mask failures
  - The downtime necessary to implement upgrades (may be zero)
  - The percentage of time the system is available outside of planned maintenance time
- **Capacity:** is the number of concurrent requests that can be handled by the service in a given time period.

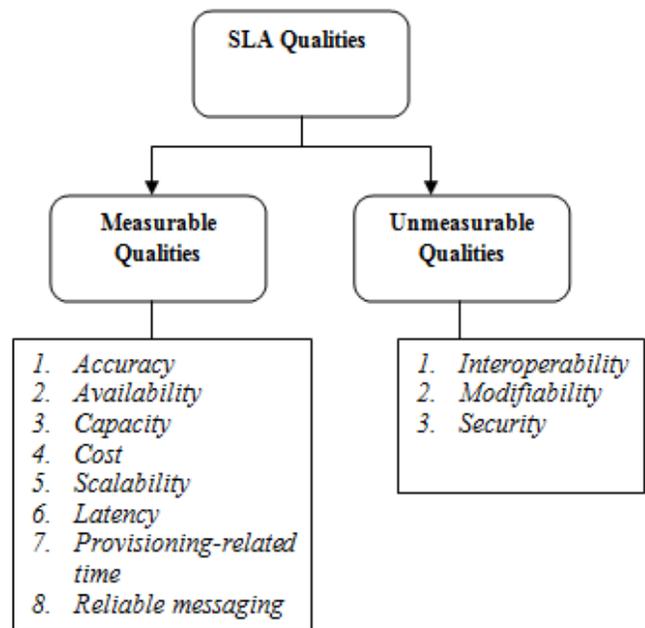


Figure (3): SLA Qualities

- **Cost:** is concerned with the cost of each service request. It is possible to specify
  - The cost per request
  - The cost based on the size of the data
  - Cost differences related to peak usage times
- **Latency:** is concerned with the maximum amount of time between the arrival of a request and the completion of that request.
- **Provisioning-related time:** (e.g., the time it takes for a new client's account to become operational)

- **Reliable messaging:** is concerned with the guarantee of message delivery. It is possible to specify
  - How message delivery is guaranteed (e.g., exactly once, at most once)
  - Whether the service supports delivering messages in the proper order
- **Scalability:** is concerned with the ability of the service to increase the number of successful operations completed over a given time period. It is possible to specify the maximum number of such operations.

**B. Unmeasurable Qualities**

Unmeasurable qualities are those that cannot be measured automatically from a given viewpoint; for example, determining the cost of changing a service is difficult to automate. Measurable qualities include interoperability, modifiability, and security.

- **Interoperability:** is concerned with the ability of a collection of communicating entities to share specific information and operate on it according to an agreed-upon operational semantics.
- **Modifiability:** is concerned with how often a service is likely to change. It is possible to specify how often the service's
  - Interface changes
  - Implementation changes
- **Security:** is concerned with the system's ability to resist unauthorized usage, while providing legitimate users with access to the service. It is also characterized as a system providing non-repudiation, confidentiality, integrity, assurance, and auditing. It is possible to specify the methods for
  - Authenticating services or users
  - Authorizing services or users
  - Encrypting the data

**V. THE PROPOSED APPROACH FOR MONITORING QoS USING SLAs IN WEB SERVICE**

An official agreement between the service provider and user is required to guarantee the defined level of the web service performance based on service quality factors. Such a service level agreement may be very comprehensive and at the same time very specific. The service level agreement may include the procedures to be followed by the provider and user in the case when either party fails to follow the agreement.

The researchers propose an approach for monitoring QoS of web service using SLA between consumer and provider as shown in Figure (4) illustrates the proposed approach for monitoring QoS using SLA. This approach used Event-based monitoring and history-based monitoring. The event-based monitoring is listening to the events in parallel to the execution of a business process to verify nonfunctional qualities of a service. History-based monitoring is an extension to event-based monitoring by collecting events in a

history event repository. It is possible to recognize upon QoS requirements that deal with a history of process executions. An example of such a requirement could be "eighty percent of the times a process goes into execution it must complete within one minute".

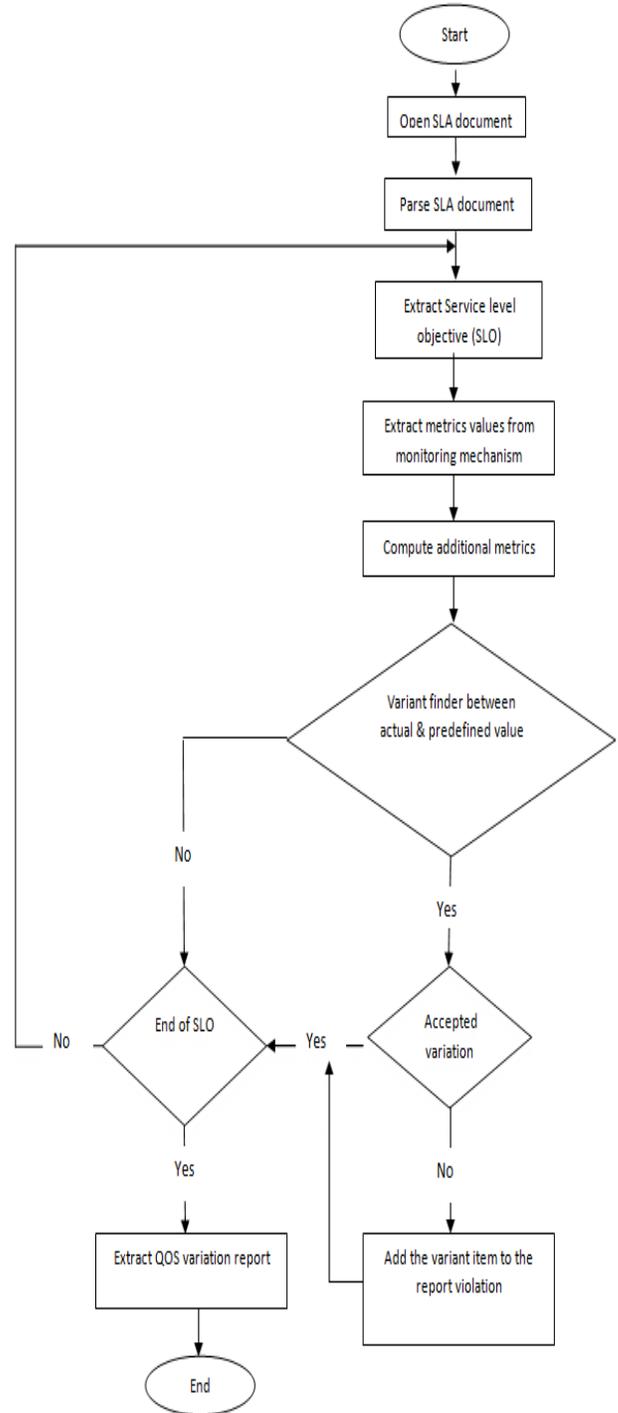


Figure (4): Proposed monitoring QoS using SLA

The proposed approach includes the following activities:

1. Open SLA document.
2. Parse SLA document.
3. Extract Service Level Objective (SLO).
4. Extract metrics values from monitoring mechanism.
5. Compute additional metrics.
6. Detect Variant finder between actual & predefined value.
7. Extract QoS variation report.

#### A. Open SLA document

In this approach, evaluation procedures run on consumer's service. The values of the SLA parameters are input for the evaluation procedure, which can run on:

- o Either the service consumer or service provider
- o Both the service consumer and service provider

#### B. Parse SLA document

As described in section, a WSLA document divides an SLA into three sections: parties, service description, and obligations as described in section 2. The WSLA language is based on XML; it is defined as an XML schema then WSLA document is parsed to save it.

#### C. Extract Service Level Objective (SLO)

In the obligations define the SLO, that is, guarantees and constraints that may be imposed on the SLA parameters. SLA parameters define the metrics values that were agreed between the service provider and the client. The following elements of the WSLA document [30] are collected and saved it to database:

- o The web service name
- o The obliged party is the name of a party that is in charge of delivering what is promised in this guarantee.
- o The start of validity period
- o The end of validity period
- o The predicate that applies for the specific value
- o The name of the SLA parameter
- o The agreed value of the SLA parameter and
- o The Evaluation Event defines when the expression of the service level objective should be evaluated.

The example shows in figure 5 a service level objective given by ACMEProvider for one month in 2002. It guarantees that the SLA parameter AverageResponseTime must be less than 0.5 if the SLA parameter Transactions is less than 10000. This condition should be evaluated each time a new value for the SLA parameter is available.

#### D. Extract metrics values from monitoring mechanism

Monitor web service to collect all the metrics that are available and composed metrics, such as minimum response time, average response, Total number of authentication failures, total number of authentication successes, total

number for requests that caused SOAP faults and Total number of successful invocations of the method. All this metrics (monitoring metrics and composed metrics) save it in database to can make History-based monitoring.

```
<ServiceLevelObjective name="slo1">
  <Obligated>ACMEProvider</Obligated>
  <Validity>
    <Start>2002-11-30T14:00:00.000-05:00</Start>
    <End>2002-12-31T14:00:00.000-05:00</End>
  </Validity>
  <Expression>
    <Implies>
      <Expression>
        <Predicate xsi:type="Less">
          <SLAParameter>Transactions</SLAParameter>
          <Value>10000</Value>
        </Predicate>
      </Expression>
      <Expression>
        <Predicate xsi:type="Less">
          <SLAParameter>AverageResponseTime</SLAParameter>
          <Value>0.5</Value>
        </Predicate>
      </Expression>
    </Implies>
  </Expression>
  <EvaluationEvent>NewValue</EvaluationEvent>
</ServiceLevelObjective>
```

Figure (5): Service Level Objective example in WSLA [36]

#### E. Compute additional metrics

There are some other metrics that can be calculated using the metrics collected from monitoring web service such as availability. Availability [37] is the probability that a service is up and running. It can be calculated in the following way:

$$\text{Availability} = (\text{MTBF} / (\text{MTBF} + \text{NTTR})) * 100$$

Mean Time Between Failures (MTBF) [37] is the average time that a web Service can perform its agreed Function without interruption. Mean Time To Repair (NTTR) [37]: The average time taken to repair a web after a Failure.

#### F. Detect Variant finder between actual & predefined value

This step is responsible for comparing between SLO (predefined value) and monitoring correspond the SLA parameters at runtime (actual value). Once the variant is detected, the variant does not fulfill the system quality requirement; the validator checks if this variant accepted or not. If the variant not accepted then the validator add this item into the report. This step is made for each SLO. In table 1, examples describe this step.

Table 1. Examples of Variant Finder Between Actual & Predefined Value

SLO Name	Constrain	Predefine metrics	Actual metric	variant accepted
Condition_SLO_for_Avgthroughput	Over utilized <0.30	Avgthroughput >1000	Avg throughput =950	Yes
Condition_SLO_for_ResponseTim	Transaction <10000	Average ResponseTime <0.5	Average ResponseTime =0.25	No Add this item to report

### G. Extract QoS variation report

After the SLO metrics is finished. The validator extracts report about the QoS variant and sent it to corrective management to carry out actions. The Management Service will retrieve the appropriate actions to correct the problem, as specified in the SLA to improve the quality of service.

## VI. CONCLUSION AND FUTURE WORK

This paper aimed to propose an approach for monitoring QoS using SLA in web Services. Therefore, the researchers studied many attempts of monitoring of web services, SLA management including QoS management, and metrics of monitoring QoS. Then, the researcher proposed an approach that consists of seven steps: open SLA document, parse SLA document, extract service level objective, extract values of metrics from monitoring mechanism, compute additional metrics, detect variant between actual & pre-defined value, and finally extract QoS variation report. The QoS variation report can help in achieving higher quality of web services. The researchers conclude that a special attention must be given to SLA in dealing with the quality of web services.

In future, the researchers aim to apply the proposed approach using real data of web services and compare the results with other results produced from other evaluation techniques of web services quality. In addition, the researchers aim to enhance the proposed approach by examining more formulas for all possible metrics, ranking and classification of the metrics to extract the most important set, and/or utilizing the good practices of other evaluation techniques.

## REFERENCES

- [1] Seong-Hoon Lee, "A Study on Web Service Analysis and Bio-information based Web Service Security Mechanism", International Journal of Security and Its Applications, Vol.8, No.2, 2014, pp.77-86.
- [2] Qusay H. Mahmoud, "Service-Oriented Architecture (SOA) and Web Services: The Road to Enterprise Application Integration (EAI)", April 2005, [online] Available: <http://www.oracle.com/technetwork/articles/javase/soa-142870.html>, [accessed 30/12/2014].
- [3] W3C, Web Services Architecture, <http://www.w3.org>, (2003).
- [4] F. Curbera, "Unraveling the Web Services Web: An Introduction to SOAP, WSDL, and UDDI", IEEE Internet Computing, Vol. 6, Issue 2, p.p. 86 – 93, August 2002.
- [5] G. Alonso, F. Casati, H. Kuno and V. Machiraju, "Web Services Concepts, Architectures and Applications Series: Data-centric Systems and Applications", Addison-Wesley, (2002).
- [6] Demian Antony D'Mello and Ananthanarayana V S, "Dynamic Web Service Composition Based on Operation Flow Semantics", International Journal of Computer Applications, Vol. 1, No. 26, p.p. 1-9, February 2010.
- [7] H. Kreger, "Web Services Conceptual Architecture (WSCA1.0)", Published May 2001, [online] Available: [www.ibm.com/software/solutions/webservices/pdf/wsca.pdf](http://www.ibm.com/software/solutions/webservices/pdf/wsca.pdf) [accessed: 2/1/2015].
- [8] Karthikeyan. J and Suresh Kumar. M, "Monitoring QoS parameters of composed web services", International Conference OnInformation Communication And Embedded Systems (ICICES), 2014.
- [9] K. Kritikos and D. Plexousakis, "Requirements for QoS-based web service description and discovery," IEEE Trans. on Service Computing, vol. 2, no. 4, pp. 320-337, October-December 2009.
- [10] Z. Balfagih and M. F. Hassan, "Quality model for web services from multi-stakeholders perspective", in Proc. Information Management and Engineering Conference, Kuala Lumpur, 2009, pp. 287-291.
- [11] M. Marzolla and R. Mirandola, "QoS analysis for web service applications: A survey of performance-oriented approaches from an architectural view point", Technical Report, February 2010.
- [12] Mohamad Ibrahim Ladan, " Web Services Metrics: A Survey and A Classification", International Conference on Network and Electronics Engineering, IPCSIT vol.11, 2011.
- [13] Katawut Kaewbanjong and Sarun Intakosum, "QoS Attributes of Web Services: A Systematic Review and Classification", Journal of Advanced Management Science, Vol. 3, No. 3, September 2014.
- [14] Shanmuga Priya R1, Kanchana R2, "AOP Based QoS Monitoring of Dynamic Web Service Compositions", IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), 2014.
- [15] Pierre Bernard, "Foundations of ITIL", Van Haren Publishing, Zaltbommel, 2012, PP. 171.
- [16] G. Dobson and A. Sanchez-Macian, "Towards Unified QoS/SLA Ontologies". Proceedings of the IEEE Services Computing Workshops (SCW 2006), 2006.
- [17] S. Ran, "A model for web services discovery with QoS", SIGecom Exchanges, 4(1):1-10, 2003.
- [18] M. Comuzzi, C. Kotsokalis, G. Spanoudkis, R. Yahyapour, "Establishing and Monitoring SLAs in Complex Service Based Systems", IEEE International Conference on Web Services 2009.
- [19] H. G. Song and K. Lee, "sPAC (Web Services Performance Analysis Center): Performance Analysis and Estimation Tool of Web Services", In Proceedings of the 3rd International Conference on Business Process Management (BPM'05), 2005, PP. 109-119.
- [20] D. Gunter, B. Tierney, B. Crowley and M. Holding, J. Lee, "Netlogger: a toolkit for distributed system performance analysis", 8th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, 2000, pp. 267-273.
- [21] T. Suzumura, T. Takase, and M. Tatsubori, "Optimizing Web services performance by differential deserialization", In Proceedings of the IEEE International Conference on Web Services(ICWS'05), 2005, pages 185-192.
- [22] B. Koller and L. Schubert, "Towards autonomous SLA management using a proxy-like approach", Multiagent Grid System, Vol.3, 2007.
- [23] M. Tian, A. Gramm, H. Ritter and J. Schiller, "Efficient Selection and Monitoring of QoS-aware Web services with the WS-QoS Framework", In Proceedings of the International Conference on Web Intelligence (WI'04), Beijing, China, 2004.
- [24] I. Brandic, D. Music, P. Leitner and S. Dustdar, "VieSLAF Framework: Enabling Adaptive and Versatile SLA Management", 6th International Workshop on Grid Economics and Business Models 2009.
- [25] F. Rosenberg, C. Platzer, and S. Dustdar, "Bootstrapping performance and dependability attributes of web service", IEEE International Conference on Web Services, 2006, pp 205-212.
- [26] L. Zeng, B. Benatallah, M. Dumas, J. Kalaganam and Q. Z. Sheng, "Quality driven web services composition", In Proceedings of the 12th

- International Conference on World WideWeb (WWW'03), pages 411–421, New York, NY, USA, 2003. ACM Press.
- [27] L. Zeng, B. Benatallah, A. H. Ngu, M. Dumas and J. Kalagnanam, and H. Chang, "Qos-aware middleware for web services composition", IEEE Transactions on Software Engineering, 30(5),p.p. 311–327, May 2004.
  - [28] A. D'Ambrogio, P. Bocciarelli, "A model-driven approach to describe and predict the performance of composite services", 6th international workshop on Software and performance, 2007, pp. 78-89.
  - [29] Service level agreement. (2015, january) wikipedia. [Online] [http://en.wikipedia.org/wiki/Service-level\\_agreement](http://en.wikipedia.org/wiki/Service-level_agreement)
  - [30] L. Baresi, C. Ghezzi and S. Guinea. "Smart Monitors for Composed Services". In Procs of ICSOC'04, New York, USA, pages 308 – 315, November 2004.
  - [31] Chrysostomos Zeginis, "Monitoring the QoS of Web Services using SLAs - Computing metrics for composed services", Master's Thesis, Heraklion, PP. 13-14, March 2009.
  - [32] E. Wustenhoff, "Service Level Agreement in the Data Center".,Sun BluePrints, April 2002.
  - [33] P. Bianco Philip, G. Lewis and P. Merson, "Service Level Agreements in Service-Oriented Architecture Environments", Technical Note of Software Engineering Institute, September 2008.
  - [34] A. Keller and H. Ludwig, "The WSLA Framework: Specifying and Monitoring of Service Level Agreements for Web Services", IBM research report RC22456, May 2002.
  - [35] R. Kassab and Aad van Moorsel, "Mapping WSLA on Reward Constructs in Mobius", In Procs of UKPEW 2008, London, England, July 2008.
  - [36] <http://www.ibm.com/developerworks/library/ws-slafram/>  
[accessed: 15/12/2015].
  - [37] <http://www.knowledgetransfer.net/dictionary/ITIL/en/Availability.htm>  
[accessed: 15/12/2015].

# Towards a Fuzzy based Framework for Effort Estimation in Agile Software Development

Atef Tayh Raslan

Department of Computer and  
Information Sciences, Institute of  
Statistical Studies and Research,  
Cairo University, Egypt

Nagy Ramadan Darwish

Department of Computer and  
Information Sciences, Institute of  
Statistical Studies and Research,  
Cairo University, Egypt

Hesham Ahmed Hefny

Department of Computer and  
Information Sciences, Institute of  
Statistical Studies and Research,  
Cairo University, Egypt

**Abstract**—Effort estimation in the domain of software development is a process of forecasting the amount of effort expressed in persons/month required to develop software. Most of the existing effort estimation techniques are suitable for traditional software projects. The nature of agile software projects is different from traditional software projects; therefore using the traditional effort estimation techniques can produce inaccurate estimation. Agile software projects require an innovated effort estimation framework to help in producing accurate estimation. The main focus of this paper is the utilization of fuzzy logic in improving the effort estimation accuracy using the user stories by characterizing inputs parameters using trapezoidal membership functions. In this paper, the researchers proposed a framework based on the fuzzy logic which receives fuzzy input parameters of Story Points (SP), Implementation Level Factor (ILF), FRiction factors (FR), and Dynamic Forces (DF) to be processed in many successive steps to produce in final the effort estimation. The researchers designed the proposed framework using MATLAB to make it ready for later experiments using real data sets.

**Keywords:** Agile software development, Effort estimation, story points, fuzzy logic

## I. INTRODUCTION

Software project management is the process of planning, organizing, staffing, monitoring, controlling and leading a software project. Project life cycle consists of four phases [9, 14]: Project initiation, project planning, project execution, and project closure. Project planning is a crucial phase in software project life cycle because it includes many challenging activities that are necessary; such as software estimation process that includes estimating the size of the software product to be produced, estimating the effort required, developing preliminary project schedules, and finally, estimating overall cost of the project.

Effort estimation is the process of predicting the amount of effort that is expressed by the number of persons per month required to develop or maintain software projects based on incomplete and uncertain requirements. Effort estimates may be used as input to project plans, iteration plans, budgets, and investment analyses.

Software development is a mentally complicated task [25]. Therefore, different software development methodologies and quality assurance methods are used in order to attain high quality, reliable, and bug free software [18]. In recent years, agile software development methods have gained much attention in the field of software engineering [32]. Agile methodologies emphasize on working software as the primary measure of progress [18, 4]. Agile methods deal with unstable and volatile requirements by using a number of techniques, focusing on collaboration between teamwork and customers and support early product delivery [4]. The traditional effort estimation techniques require modifications or improvements to be suitable for agile software methodologies. Therefore, this paper aims to propose an innovated effort estimation framework to help in producing accurate estimation. The proposed framework depends on utilization of fuzzy logic, SP, ILF, FR, and DF.

The rest of this paper is divided into seven sections. Section II introduces an overview on agile software development that includes main characteristics of agile software process and examples of agile methods. Section III presents a brief introduction of fuzzy logic. Section IV introduces a literature review that includes some important papers and researches in effort estimation. Section V introduces effort estimation techniques and focuses on story points. Section VI introduces the proposed framework and explains the main steps required to reach a final effort estimation using this framework. Section VII introduces the conclusion of the paper including the main ideas discussed in the paper. Section VIII introduces the ideas that are expected to be focused in the future.

## II. AGILE SOFTWARE DEVELOPMENT

Agile Software Development (ASD) is a group of software development processes that are iterative, incremental, self-organizing and emergent [10]. In addition, it can be defined as a connotation of flexibility, nimbleness, readiness for motion, activity, dexterity in motion, and adjustability [17]. Agile methodologies are a lightweight, efficient, low-risk, flexible, predictable, scientific, and fun way to develop software. In ASD, the time and resources are available which is generally

considered to be fixed [17]. In traditional methods, the time and resources are flexible while the functionality is considered fixed in comparison with the agile one [17]. The agile Manifesto was created in February 2001 that included the twelve principles on agile software development [7]. There are several agile methods based on the idea of agile manifesto. Examples of such methods are Dynamic Systems Development Method (DSDM), eXtreme Programming (XP), Feature Driven Development (FDD) and SCRUM [4, 10, 13, 17, 19, 24, 29].

The characteristics of ASD process include: modularity on development process level, iterative with short cycles, time-bound with iteration cycles, economic in development process, adaptive, incremental, minimize the risks, people oriented, and collaborative and communicative [10,17]. The advantages of agile software development include: Revenue, quality, visibility, risk management, agility, customer satisfaction and help to generate the right product. The revenue refers to the incremental nature of agile development enabling to be realized early as the product continues to develop. The visibility refers to encourage throughout the product development and a very cooperative approach. In addition, in ASD the small incremental releases are visible to the product owner and product team which help them to identify any issues early and make it easier to respond to change.

In traditional methodologies, a team member’s workload capacity is determined by the manager who estimates how long certain tasks will take and then assigns work based on that team member’s total available time. On other hand, in agile methodology the works are assigned to an entire team, not to an individual member. In addition, agile methodology adopts self-organization of team as a success factor that must be encouraged. Figure 1 shows the effort estimation in scrum methodology which takes place in the pre-game phase. In the sprint planning meeting, the team sits down to estimate its effort for the stories in the backlog. The team shares the concepts and scales that were learned. The product owner needs these estimates to prioritize items in the backlog, and forecast releases based on the team’s velocity [33].

### III. FUZZY LOGIC

Fuzzy Logic (FL) is a methodology to solve problems which are too complex to be understood quantitatively [5]. It is based on fuzzy set theory and introduced in 1965 by Lotfy Zadeh [1]. The Fuzzy Logic System deals with fuzzy parameters, which address imprecision and uncertainties, by mapping out the path of a given input to an output using the computing framework called the Fuzzy Inference System (FIS). A fuzzy Inference system is a knowledge based or rule based system. FIS consists of four components they are: Fuzzifier, Fuzzy Rule Base, Fuzzy Inference Engine, and Defuzzification. The fuzzifier converts the crisp input into a fuzzy set that is a non-traditional type of sets which allows an element to have a partial degree of membership. The membership refers to the degree of inclusion to specific set .There is many membership

functions but in this research we will use the triangle membership function is represented by Eq. (1) in below [12]:

$$\text{Triangle}(x: a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right) \quad (1)$$

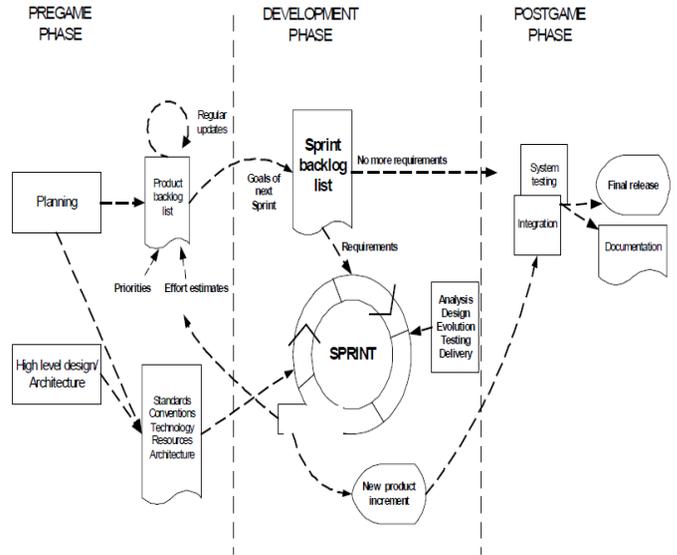
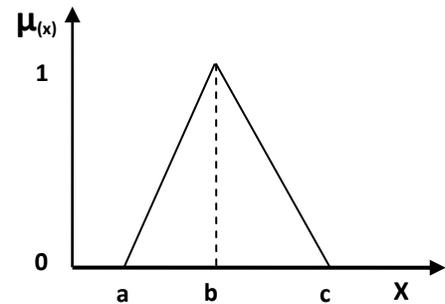


Figure 1: Scrum process [29]

Figure 2 shows the shape of the triangular membership function.



$$\text{Triangle}(x: a, b, c) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ \frac{c-x}{c-b} & b \leq x \leq c \\ 0 & x > c \end{cases}$$

Figure 2: Triangular membership function[12]

A fuzzy inference engine is a collection of IF -THEN rules stored in fuzzy rule base is known as inference engine. Defuzzification is the process which refers to the transform of fuzzy output into crisp output. The most common tools that is used for fuzzy systems is MATLAB which used for defining the input, output, inference rules, and the shape of membership function for the fuzzy system.

Software effort estimation is one of the most challenging activities in project development phases [22, 34]. However, the process of effort estimation is uncertain in nature as it largely depends upon some attributes that are quite unclear during the early stages of development. Fuzzy logic provides the concept of fuzzy sets to handle vague and inaccurate data [1].

#### IV. LITERATURE REVIEW

There are many papers and researches in effort estimation. The following are some examples of these literatures:

- Masateru T., et al, introduced revisiting software development effort estimation based on early phase development activities. This research aims to examine the relationship of early phase effort and software size with software development effort. The proposed model was constructed using early phase effort and software size. This model was evaluated by using the International Software Benchmarking Standards Group (ISBSG) dataset that collected from software development companies. The result of the experiment showed that when both software size and sum of effort needed for planning and requirement analysis phase were used as explanatory variables, the estimation accuracy was most improved [23].
- Andreas S., et al, introduced a guide to estimates the effort in agile software development. This study provides an investigation about estimation possibilities, especially for XP. It is focused on characteristics of agile methodologies. This study provides guidelines for measurement aspects within XP project. The proposed guide was evaluated using a survey which consists of 17 questions. Each question considers reflect the agile methodologies and effort estimation. The survey shows that, the benefit of agile methods is difficult to assess at the moment. Especially the costs of maintenance projects must be taken into account for it [3].
- Amrita R., et al, introduced a guide to optimizing the cost and effort in scrum projects using function point with COCOMO model. The proposed guide used on the real case study Xsset for estimation. Furthermore, it is show the different tracking tools and representation used to optimize the performance like run chart, JIRA, SVN and SCTM. This research was concluded the agile methodology is very effective in software development. Also, they concluded that by merging the two estimation models can be improving the effort applied and saving the cost [2].
- Shahrukh Z., et al, introduced an effort estimation model for agile software development to accommodate most of the characteristics of agile methodology, especially adaption and iteration, where it is focused on user stories of as base for estimation. The model was evaluated using the empirical data collected from 21 software projects. The experimental results show that model has good estimation accuracy in terms of the Mean Magnitude of Relative Error (MMRE) and probability of a project having a relative error of less than or equal to L PRED (L). MMRE and PRED are computed from the relative error, which is the relative size of the difference between the actual and estimated value of individual effort [33].
- Ziauddin A., et al, introduced a fuzzy logic based software cost estimation Model. This study aims to utilize a fuzzy logic model to improve the accuracy of software effort estimation. The main idea in this study is fuzzifying input parameters of COCOMO II model and the result is defuzzified to get the resultant Effort. Triangular fuzzy numbers are used to represent the linguistic terms in COCOMO II model. The results of this model are compared with COCOMO II and Alaa Sheta Model. The proposed model yields better results in terms of MMRE, PRED (n) and Variance Account for (VAF). VAF is used in the context of statistical models whose main purpose is the prediction of future outcomes on the basis of other related information [34].
- Vishal S., et al, introduced optimized fuzzy logic based framework for effort estimation in software development. The performance of the proposed framework is demonstrated in terms of empirical validation carried on live project data of the COCOMO public database. The performance of the framework is evaluated using live project data of the COCOMO public database. It is shows that the proposed framework can be deployed on COCOMO II environment with experts providing required information for developing fuzzy sets and an appropriate rule base [30].
- Ratnesh L. and Abhay K., introduced an estimation model for agile web based that aim to address the numerous open issues in web development particularly in context to agile software development by analyzing data from completed agile Web development projects. This study identifies the difference between traditional and web projects. Furthermore, it is offers a tool for effort estimation based on agile manifestoes and web characteristics [26].
- Abeer H. introduced a model based on a fuzzy logic for enhancing the sensitivity of COCOMO cost model. This

study enhances the accuracy and sensitivity of COCOMO 81 intermediate by fuzzifying the cost drivers. This model was implemented through the MATLAB. The dataset was collected from six NASA centers and covers a wide range of software domains, development process, languages and complexity, as well as fundamental differences in culture and business practices between each centre. The results showed that the sensitivity of the proposed fuzzy model is superior to COCOMO81 intermediate [1].

- Jitender Choudharia and Ugrasen Suman, introduced software Maintenance Effort Estimation Model (SMEEM) for software maintenance estimation. The proposed model uses SP to calculate the volume of maintenance and value adjustment factors that are affecting story points for effort estimation. The proposed model is illustrated with various types of maintenance projects such as web based application, MIS and critical project. It is designed to help the project manager in charge of software maintenance to calculate the estimated software maintenance effort in terms of Adjusted Story Point (ASP), size, cost and duration. It generates the more realistic and precise estimation results. This model is applicable only for agile and extreme programming based maintenance environment [16].

#### V. EFFORT ESTIMATION TECHNIQUES

Software effort estimation techniques can be classified into algorithmic and non-algorithmic models [26]. The most popular models that were classified as non-algorithmic technique are:

- *Expert judgment.*
- *Thumbs rule.*
- *Delphi technique.*
- *Wide band delphi technique.*
- *Parkinson's law.*
- *Pricing to win.*
- *Probe software process.*
- *Team software process technique planning.*

The recent researches focused on an algorithmic model such as [6, 20, 26]:

- *Line Of Code (LOC).*
- *Functions point matrices.*
- *COCOMO and COCOMO-II.*
- *Object points.*
- *Software life cycle management.*
- *Use case estimation.*
- *Story points [8,11,33].*

Algorithmic models are based on the statistical analysis of historical data. These models require accurate input of specific attribute such as line of code LOC, function points FP, number of user screens, complexity, and velocity [8, 33].

In this paper, the researchers focused on the story point for estimating the efforts in agile software development. The story point is the most common estimation technique for agile software development [8, 11, 33].

Story points refer to an estimate of the relative scale of the work in terms of actual development effort. Story points are usually expressed either in numbers that follow the Fibonacci series, t-shirt sizes (XS, S, M, L, XL, XXL) or even dog breeds (Chihuahua to Great Dane) [33]. Fibonacci series are a sequence of numbers where each number is the sum of the previous two. Fibonacci numbers are 1, 2, 3, 5, 8, 13, 21, etc. Story point estimation is done using relative sizing by comparing one story with a sample set of previously sized stories. Relative sizing across stories tends to be much more accurate over a larger sample, than trying to estimate each individual story for the effort involved. Planning poker is one of the most useful tools in agile software development [31].

In ASD, each member of the team is given a group of index cards. Each index card represents a size in user story sequence. Each member of the team then chooses a card representing their estimate of the effort involved in completing a particular user story, and places that card face down in front of them.

The agile effort estimation includes story size, complexity, implementation level factor, and velocity [8]. Story size refers to an estimate of the relative scale of the work in terms of actual development effort. The values assigned are usually taken from the Fibonacci sequence. The assigned values can be changed by the team itself or even the criteria can be redefined. The Complexity (Com) scale assigned like a story size by using Fibonacci sequence. The user stories grouped according to the characteristics of agile methods [21]. Also, these groups can be adjusted by the team itself.

The velocity refers to how much product backlog effort a team can handle in one unit of time. Actual velocity of any development project varies with the size and experience of a team [34]. Computing velocity helps the team to improve their estimates over the life span of a project. In addition, the velocity refers to the team's capacity which enables release planning and over time calculations.

Optimization process refers to studying the project constraints which should be completed before the calibration to improve the stability of the velocity calculation. This process includes two factors they are FR and DF. The friction includes a range of factors that may affect the team velocity they are team composition, process, environmental factors, and team dynamic [15]. The composition of the team concerns the skills and attitudes of team members. The process factor refers to the percentage of change in the agile methods, release, building and testing. The environmental factors refer to the noise, poor ventilation, poor lighting, uncomfortable seating and desks, inadequate hardware or software. Team dynamics are patterns of interaction among team members that determine the performance of the team.

Table 1 shows the FR with a range of values. Each factor has been adjusted according to their risk level (stable, volatile, highly volatile, and very highly volatile).

TABLE 1: FRICTION FACTORS [33]

Friction factors	Stable	volatile	Highly volatile	Very highly volatile
Team composition	1	.98	.95	.91
Process	1	.98	.94	.89
Environmental factors	1	.99	.98	.96
Team dynamic	1	.98	.91	.85

DF refers to the factors that could lead to loss of velocity [33]. These factors are unpredictable and unexpected. DF including nine factors: team changes, new tools, vendor defects, responsibilities out of the project, personal issues, stakeholders, unclear requirements, changing requirements, and reallocation. Those factors are ranked from (Normal, high, very high, and extra high). Table 2 shows the assigned values for those factors.

TABLE 2: VARIABLE FACTORS [33]

Variable factors	Normal	High	Very High	Extra High
Expected to team change	1	.98	.95	.91
Introduction to a new tools	1	.99	.97	.96
Vendor's Defect	1	.98	.94	.90
Team member's responsibilities outside the project	1	.99	.98	.98
Personal Issues	1	.99	.99	.98
Expected Delay in Stakeholder response	1	.99	.98	.96
Expected Ambiguity in Details	1	.98	.97	.95
Expected Changes in environment	1	.99	.98	.97
Expected Relocation	1	.99	.99	.98

## VI. PROPOSED FRAMEWORK

The process of effort estimation is uncertain in nature as it largely depends upon some attributes that are quite unclear during the early stages of development [26]. Fuzzy logic provides the concept of fuzzy sets to handle vague and inaccurate data. Its address imprecision and uncertainties, by mapping out the path of a given input to an output using the computing framework called the FIS. It consists of four main components; they are: Fuzzifier, fuzzy rule base, fuzzy inference engine, and defuzzification. The fuzzifier converts the crisp input into a fuzzy set by using a membership function. Fuzzy rule base which represented by if-then rules. Fuzzy

inference engine is a collection of if-then rules. Defuzzification is the process that refers to the translation of fuzzy output into crisp output. The MATLAB Fuzzy Inference System was used in the fuzzy calculations.

Figure 3 shows that the flowchart of the proposed framework starts with inputs *SP, ILF, FR, and DF*. Then, these inputs are fuzzified to obtain fuzzy sets using the triangular member function. Rule base includes a group of IF-Then rules that define the complexity and velocity fuzzy variables. The FIS evaluates all rules of the rule base. The resulted fuzzy set is deuzzified to a crisp value and then the complexity and velocity are calculated. Finally, the effort is estimated using the resulted complexity and velocity.

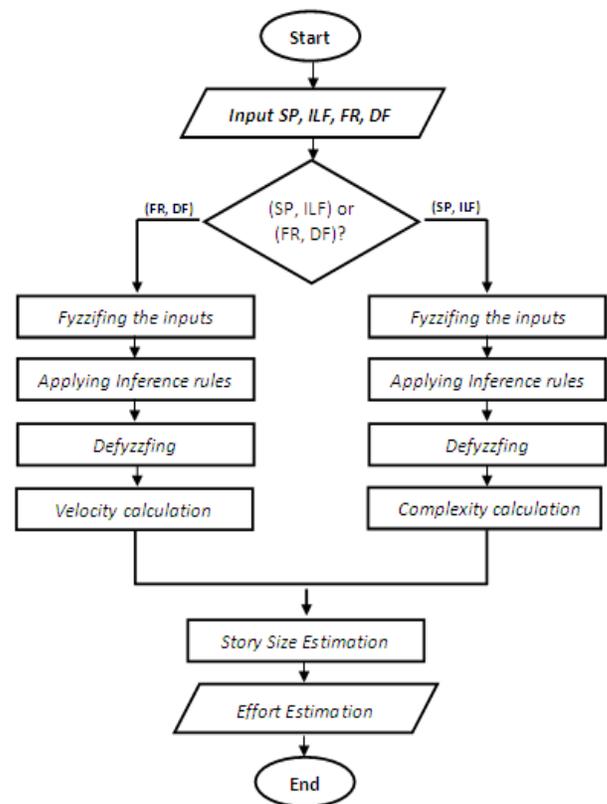


Figure 3: Proposed framework Flowchart

Figure 4 shows the proposed framework for effort estimation in agile project based on fuzzy logic which consists of three phases: The first phase refers to input that includes four inputs *SP, ILF, FR, and DF*. These input variables are changed to fuzzy variables based on fuzzification process. In the story point the terms Very Small Story (VSS), Small Story (SS), Medium Story (MS), Large Story (LS), and Very Large Story (VLS) were defined for that input.

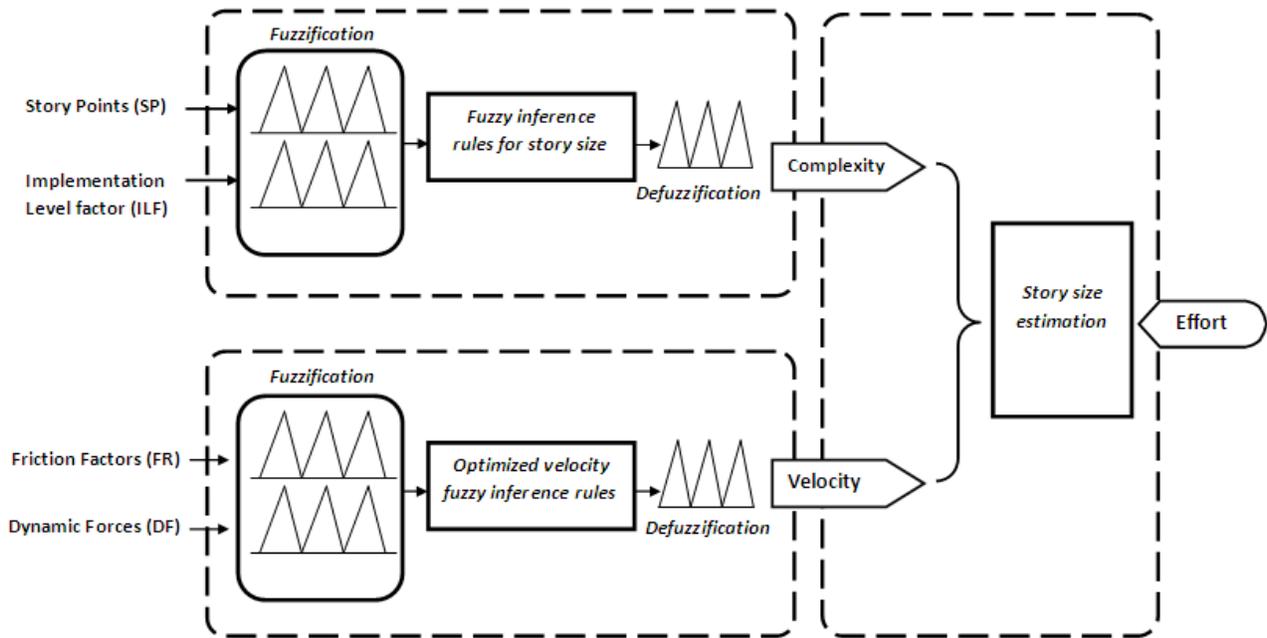


Figure 4: Proposed Framework for Effort Estimation ASD

Table 3 shows the weight of SP which can have many iterations depending upon the system requirements. The weights assigned for each user stories are predicted by using Fibonacci sequence.

TABLE 3: SP WEIGHTS

Term	Wight
Very Small Story (VSS)	1
Small Story (SS)	2
Medium Story (MS)	3
Large Story (LS)	5
Very Large Story (VLS)	8

Figure 5 shows the linguistic variables for a story point which has a range of weight (VSS, SS, MS, LS, and VLS).

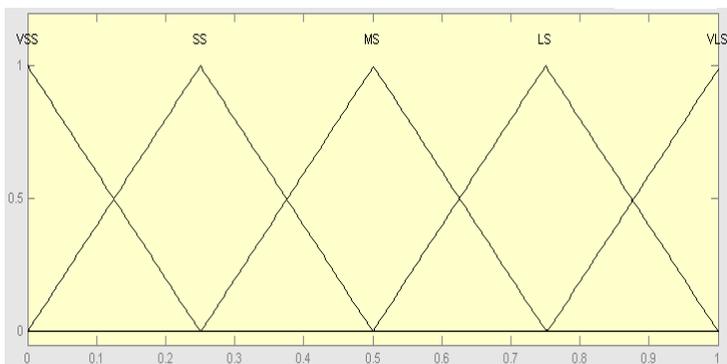


Figure 5: Linguistic variables for SP

In ILF includes the terms Off the Shelf (OS), Full Experience Components (FEC), Partial Experience Components (PEC), and New Components (NC). Table 4 shows the scaling factor for implementation level which represents the level of understanding each user story.

TABLE 4: ILF SCALING

Term	Scale
Off the Shelf (OS)	1
Full Experience Components (FEC)	2
Partial Experience Components (PEC)	3
New Components (NC)	4

Figure 6 shows the linguistic variables for implementation level which has a range of weight (OS, FEC, PEC, and NC).

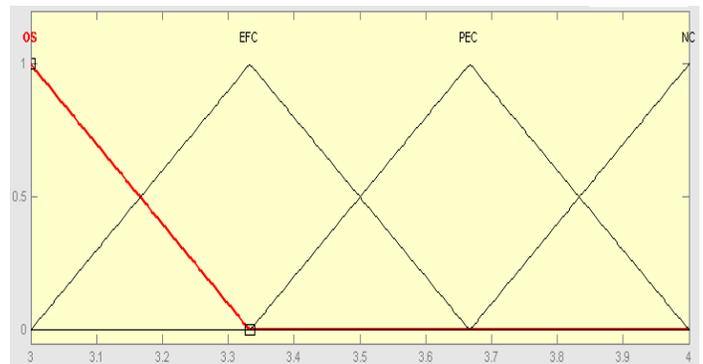


Figure 6: Linguistic variables for ILF

Table 5 shows the Com in the terms Very Low (VL), Low (L), Medium (M), High (H), and Very high (VH). The complexity has been determined based on the user stories in ILF. For example, if the user story is very low and ILF is OS, then Com is very low.

TABLE 5: COM LEVELS

Term	Level
Very Small Story (VSS)	1
Small Story (SS)	2
Medium Story (MS)	3
Large Story (LS)	5
Very Large Story (VLS)	8

Figure 7 shows the linguistic variables for a complexity which has a range of weight (VL, L, M, H, and VH).

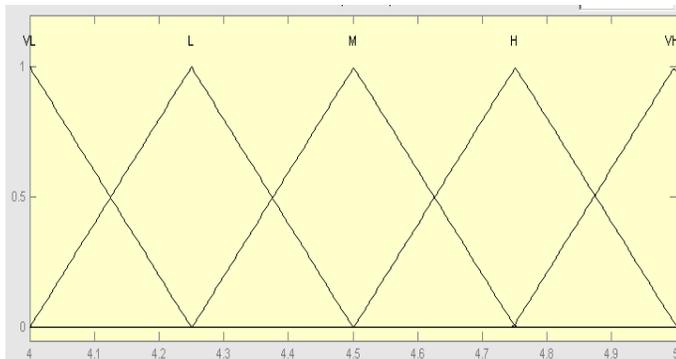


Figure 7: Linguistic variables for Com

The effort for the complete project will be the sum of membership degree of SP, ILF, and Com for all individual user stories. The Eq. (2) show the fuzzy story size (FSS) by using the triangle membership function  $\mu$ .

$$FSS = SP \cdot \mu_{SP} \times \sum_i^4 ILF \cdot \mu_i^{ILF} \times \sum_i^5 Com \cdot \mu_i^{Com} \quad (2)$$

FR includes the terms Stable(S), Volatile (V), Highly Volatile (HV), and Very Highly Volatile (VHV) were defined for the four variables team composition, process, environmental factors, and team dynamic. Figure 8 shows the linguistic variables for a team composition which has a range of weight (S, V, HV, and VHV).

DF includes the terms Normal (N), High (H), Very High (VH), and Extra High (EH) were defined for the nine variables they are: Expected to team change, introduction to new tools, vendor's defect, team member's responsibility outside the project, personal issues, expected delay in stakeholder response, expected ambiguity in details, expected changes in environment, and expected relocation. Figure 9 shows the

linguistic variables for expected to team change which has a range of weight (N, H, VH, and EH).

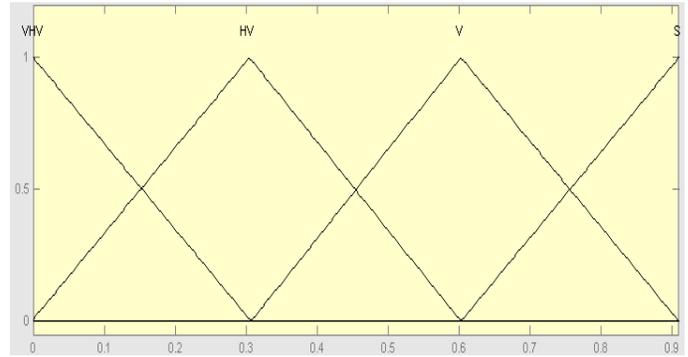


Figure 8: Linguistic variables for FR

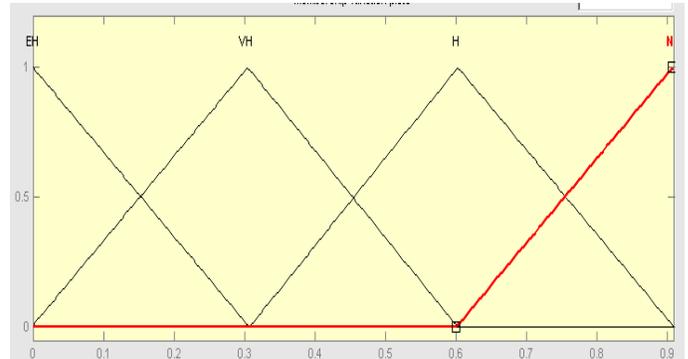


Figure 9: Linguistic variables for DF

Fuzzy FR (FFR) is calculated as product of all four components of FR and the membership degree. Eq.3 Shows the FFR calculation by using the triangle membership function  $\mu$ :

$$FFR = \prod_{i=1}^4 FR \cdot \mu_i^{FF} \quad (3)$$

The Deceleration (D) represents the rate of negative change of velocity. The D and adjusted Velocity (V) are calculated as [33]:

$$D = FR \times DF \quad (4)$$

$$V = (VI)^D \quad (5)$$

The compilation time (T) is calculates by the following formula [33]:

$$T = \frac{E}{V} \quad (6)$$

The second phase is inference engine which run on a set of fuzzy rules. The Fuzzy Model rules contain the linguistic variables related to the agile project. This is sample of the rules that represent the proposed framework below:

*IF SP is VSS and ILF is OS THEN Com is VL*  
*IF SP is VLS and ILF is NC THEN Com is VLS*  
*IF SP is MS and ILF is OS THEN Com is SS*  
*IF SP is LS and ILF is PEC THEN Com VLS*

The third phase a defuzzification process which is calculates and converts the fuzzy output into crisp data form. The most common method is a Centeroide Method or it is called Center Of Area (COA) [34].

$$COA = \frac{\sum_{x=a}^B \mu(x).X}{\sum_{x=a}^B \mu(x)}$$

## VII. CONCLUSION

This paper aimed to propose a fuzzy based framework for effort estimation in agile software development. Therefore, the researchers studied many researches in the domain of effort estimation of agile software development either based on fuzzy logic or not. In addition, the researchers studied in detail the features of agile software development, story points, and fuzzy logic. Then, the researchers proposed a framework that is based on the fuzzy logic which receives fuzzy input parameters of story points, implementation level factor, friction factors, and dynamic forces. The proposed framework starts with inputs *SP*, *ILF*, *FR*, and *DF*. Then, these inputs are fuzzified to obtain fuzzy sets using the triangular member function. Rule base includes a group of IF-Then rules that define the complexity and velocity fuzzy variables. The fuzzy inference system evaluates all rules of the rule base. The resulted fuzzy set is deuzzified to a crisp value and then the complexity and velocity is calculated. Finally, the effort is estimated using the resulted complexity and velocity.

The researchers thought that the utilization of fuzzy logic and story point in the proposed framework may help in producing an accurate estimation of effort. The researchers designed the proposed framework using MATLAB to make it ready for later experiments using real data sets.

## VIII. FUTURE WORK

The ideas that are expected to be focused in the future include:

- Applying the proposed framework in this paper using real data set from software projects and comparing the results with other results produced from different techniques.
- Enhancing the framework using the results of the application.

- Utilization of COCOMOII to enhance the proposed framework

## REFERENCES

- [1] Abeer Hamdy, "Fuzzy Logic for Enhancing the Sensitivity of COCOMO Cost Model", Journal of Emerging Trends in Computing and Information Sciences, Vol.3, No.9, 2012.
- [2] Amrita Raj Mukker, Anil Kumar Mishra, and Latika Singh, "Enhancing Quality in Scrum Software Projects", International Journal of Science and Research (IJSR), Vol. 3, Issue 4, 2014.
- [3] Andreas Schmietendorf, Martin Kunz, and Reiner Dumke, "Effort estimation for Agile Software Development Projects", 5th Software Measurement European Forum, 2008.
- [4] Ann L. Fruhling and Alvin E. Tarrell, "Best Practices for Implementing Agile Methods", IBM, 2008.
- [5] Anupama Kaushik, A.K. Soni, and Rachna Soni, "A Type-2 Fuzzy Logic Based Framework for Function Points",
- [6] Ashita Malik, Varun Pandey, Anupama Kaushik, "An Analysis of Fuzzy Approaches for COCOMO II", International Journal of Intelligent Systems and Applications, Vol.5, 2013.
- [7] Beck et al., "http://agilemanifesto.org/iso/en/manifesto.html," last visited in December 2014.
- [8] C. Sathish Kumar, A. Anitha Kumari, and R. Srinivasa Perumal, "An Optimized Agile Estimation Plan Using Harmony Search Algorithm," International Journal of Engineering and Technology (IJET), Vol 6 No 5, 2014.
- [9] Daniele Di Filippo, Russell D. Archibald, and Ivano Di Filippo, "Six-Phase Comprehensive Project Life Cycle Model", Vol. I, Issue V, December 2012.
- [10] David Cohen, Mikael Lindvall, Patricia Costa, "An Introduction to Agile Methods," Advances in computers, Elsevier, V. 62, pp. 20-22, 2004.
- [11] Evita Coelho, Anirban Basu, "Effort Estimation in Agile Software Development using Story Points," International Journal of Applied Information Systems (IAIS), Volume 3- No.7, August 2012.
- [12] Ganesh M., "Introduction to Fuzzy Sets and Fuzzy Logic", Prentice-Hall, 978-8120328617, 2006.
- [13] Ioannis G. Stamelos, and Panagiotis Sfetos, "Agile software development quality assurance", Idea Group, ISBN 978-1-59904-216-9, 2007.
- [14] Jason Westland, "Project management guidebook", Method123, 2003.
- [15] Jeff Dyer, Jeffrey H. Dyer, William G., "Team Building: Proven Strategies for Improving Team Performance", Jossey-Bass, ISBN: 978-1118416143, 2013.
- [16] Jitender Choudharia and Ugrasen Suman, "Story Points Based Effort Estimation Model for Software Maintenance", Elsevier, Procedia Technology, Volume 4, 2012.
- [17] John Hunt, "Agile software construction," Springer, ISBN-10: 1-85233-944-6, 2006.
- [18] K.Usha, N.Poonguzhali, and E.Kavitha, "A Quantitative Approach for Evaluating the Effectiveness of Refactoring in Software Development Process", International Conference on Methods and Models in Computer Science, Delhi, India, Dec. 2009.
- [19] Kevin Vlaanderen, Sjaak Brinkkemper, Slinger Jansen, Erik Jaspers, "Applying SCRUM Principles to Software Product Management," Information and Software Technology, V. 53, Issue 1, pp. 58-70, 2011.
- [20] Komal Garg, Paramjeet Kaur, Shalini Kapoor, and Shilpa Narula, "Enhancement in COCOMO Model Using Function Point Analysis to Increase Effort Estimation", IJCSMC, Vol. 3, Issue. 6, 2014.
- [21] Luigi Buglione, and Alain Abran, "Improving Estimations in Agile Projects: Issues and Avenues", Software Measurement European Forum (SMEF), 2007.
- [22] M.C. Ohlsson, C. Wohlin, and B. Regnell, "A project effort estimation study", Elsevier Science, information and software technology, Vol.40, 831- 839, 1998.

- [23] Masateru Tsunoda, Koji Toda, and Kyohei Fushida, Yasutaka Kamei, Meiyappan Nagappan, and Naoyasu Ubayashi, "Revisiting Software Development Effort Estimation Based on Early Phase Development Activities", 10th conference on mining software repositories (MSR), IEEE, 2013.
- [24] Mike Cohn, "Software Development Using Scrum," Addison-Wesley, ISBN: 978-0321579362, 2009.
- [25] Nagy Ramadan Darwish, "Improving the Quality of Applying eXtreme Programming (XP) Approach", International Journal of Computer Science and Information Security (IJCSIS) – ISSN 1947-5500, Vol. 9 No. 11, November 2011.
- [26] Ratnesh Litoriya and Abhay Kothari "An Efficient Approach for Agile Web Based Project Estimation: AgileMOW", Journal of Software Engineering and Applications, VOL.6, 2013.
- [27] Sandeep Kad, Vinay Chopra, "Fuzzy Logic based framework for Software Development Effort Estimation", An International Journal of Engineering Sciences (IJES), Vol. 1, 2011.
- [28] Siva Dorairaj, James Noble, and Petra Malik, "Understanding Team Dynamics in Distributed Agile Software Development", Springer, ISBN: 978-3-642-30350-0, 2012.
- [29] Warsta J., Ronkainen J., Salo O., Abrahamsson P., "agile software development", VTT, ISBN 951-38-6009-4, 2002.
- [30] Vishal Sharma and Harsh Kumar Verma, "Optimized Fuzzy Logic Based Framework for Effort Estimation in Software Development", International Journal of Computer Science Issues (IJCSI), Vol. 7, Issue 2, No 2, 2010.
- [31] Vitaliy Zurian, "http://www.agilemarketing.net/winning-planning-poker/", last visited on June 2015.
- [32] Yang Yong and Bosheng Zhou, "Evaluating Extreme Programming Effect through System Dynamics Modeling", International Conference on Computational Intelligence and Software Engineering (CiSE), Wuhan, China, Dec. 2009.
- [33] Ziauddin Zia, Shahid Kamal Tipu, and Shahrukh Zia "An Effort Estimation Model for Agile Software Development", Advances in Computer Science and its Applications (ACSA), Vol.2, 2012.
- [34] Ziauddin, Shahid Kamal, Shafiqullah Khan and Jamal Abdul Nasir, "A Fuzzy Logic Based Software Cost Estimation Model", International Journal of Software Engineering and Its Applications (IJSEIA), Vol. 7, Issue 2, 2013.

# A Novel architecture for improving performance under virtualized environments

A.P. Nirmala

Research Scholar, Karpagam University, Coimbatore &  
Assistant Professor, New Horizon College of Engg.,  
Bangalore, India

Dr. R. Sridaran

Dean, Faculty of Computer Applications,  
Marwadi Education Foundation's Group of Institutions,  
Rajkot, India

**Abstract**—Even though virtualization provides a lot of advantages in cloud computing, it does not provide effective performance isolation between the virtualization machines. In other words, the performance may get affected due the interferences caused by co-virtual machines. This can be achieved by the proper management of resource allocations between the Virtual Machines running simultaneously. This paper aims at providing a proposed novel architecture that is based on Fast Genetic K-means++ algorithm and test results show positive improvements in terms of performance improvements over a similar existing approach.

**Keywords**- Virtualization, Performance, Performance Interference, Scheduling Algorithm, Throughput

## I. INTRODUCTION

In the recent era, virtualization technology provides the advantages in the form of better manageability, optimistic provisioning and minimizing the cost in current cloud computing environments.

Virtualization allows sharing of server resources on-demand thereby creating new business opportunities. This leads to developments of new delivery models for a wider set of enterprise services. Thus, virtualization is a key enabling factor not only for Cloud Computing but also for utility computing paradigm [2][16]. However, virtualization may also lead to the contention of shared resources on each platform between virtual machines (VMs) involved which needs to be addressed.

Virtualization technology enables diverse applications to run in the isolated environments by creating multiple VMs on a single physical machine and managing resource sharing across VMs by virtual machine monitor (VMM) technology.

VMMs or hypervisors from VMware™, Xen™ community, Microsoft™ and others manage the VMs running on a single platform and ensure that they are functionally isolated from one another as shown in Fig. 1.1 [5][16].

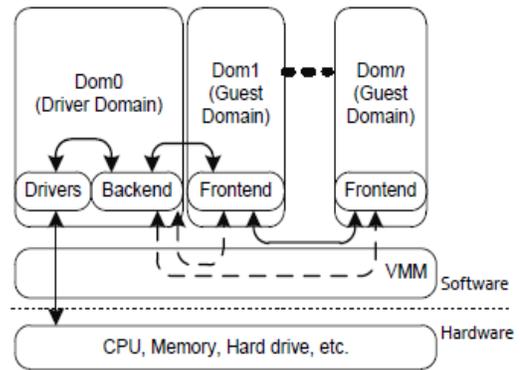


Fig. 1.1 Virtualized cloud environment.

The VMM is responsible for allocating basic resources such as CPU cycles, memory capacity, disk and network I/O bandwidth. At a high level, allocating a specific share of physical resources to a VM results in a specific performance that is measurable using certain performance metrics such as response time and/or throughput. However, as all VMs share the same physical resources, they also mutually influence each other's performance. According to X. Pu et al. [5], VMMs (hypervisors) have the abilities to slice down resources and allocate the shares to different VMs where the applications running on one VM may still affect the performance of applications running on its neighborhood VMs.

However virtualization provides features such as security, fault and environment isolations, it does not help in offering performance isolation between VMs in an effective manner. Y. Koh et al. [3], describes that a user running the same VM that belongs to the hardware but at different times will realize wide disparity in performance based on the work carried out on other VMs on that physical host.

It is essential to develop architectural techniques that ensure appropriate sharing of resources allocated to VMs running simultaneously based on their importance or their behavior. Effective management of virtualized cloud environments introduces new and unique challenges, such as efficient CPU scheduling for VMs, effective allocation of VMs to handle both CPU intensive and I/O intensive workloads. Based on this,

many novel scheduling algorithms can be thought of. These algorithms may have their primary objectives as either of minimizing the negative impacts of co-located applications or improving the overall system performance [4][1]. Ron C. Chiang et al. in [1] discusses about how the system can make optimized scheduling decisions that lead to significant improvements in both application performance and resource utilization.

According to Ron C. Chiang et al. in [1], K-means algorithm neither guarantees to converge to a global minimum nor to achieve the best optimal solution available. Genetic Algorithm (GA) is used for the purpose of finding the global minima [14]. Ron C. Chiang et al. [1] proposed the K-means++ algorithm to make a good choice of initial k centers to replace K-means algorithm that is designed to pick points that are far away from each other. Instead of choosing the point farthest from chosen points, k-means++ pick each point at random with probability proportional to the squared distance. Thus k-means++ is combined with genetic algorithm to find optimized solution.

This motivates us to propose Fast genetic k-means++, a scheduling algorithm to improve the performance in the virtualized environments. It is implemented by conducting a comprehensive evaluation with a variety of cloud applications to measure the performance in terms of application throughput, runtime and cost.

Fast genetic k-means++ algorithm, aims at improving the performance by scheduling the task to various VMs with minimized interference effects from co-located applications. It helps to reduce the runtime and improve the I/O throughput for data-intensive applications in a virtualized environment. When the task arrives, the scheduler proceeds with generating number of possible assignments of the same based on the incoming tasks and list of available VMs. Then the scheduler makes the decision for scheduling and assigning the task to different servers based on the predictions.

This paper is organized as follows, section 2 provides the related work, section 3 explains the proposed methodology and experimental results are discussed in section 4. Section 5 gives the conclusion with future work..

## II. RELATED WORK

X. Pu et al.. [5] focus on performance interference among VMs running the network workloads in virtualized environments. In their work, system-level characteristics are considered as metrics to identify the impact of running different combinations of workloads of different file sizes on the aggregate throughput. Extensive experiments are conducted to compare and understand better the combination of different workloads and the multiple factors that may cause performance interferences. However, the above work does not describe how to mitigate the I/O interference for data-intensive applications. There are several studies to evaluate the

performance degradation of VMs due to interference which are illustrated in [7][3]. Although these studies aim at proposing different types of benchmarks to identify the VM interference but they do not explain how to mitigate the interference effects.

D. Novakovic et al. proposed DeepDive, a system for VM migration [19]. DeepDive identifies interference-inducing VM and shifts VM to destination physical machine on which least interferences are reported. In contrast to DeepDive, in our proposed work, placement of task in appropriate VM alone is considered rather than VM placement itself in order to reduce the interferences.

Recently, Paragon [20], proposed test benchmarks to identify sources of interferences and their impact on co-located applications. Paragon uses previously scheduled applications to identify the best placement for new application with respect to interference in place of profiling. Our proposed work is analogues to Paragon [20] as for as the placement of application is concerned. However, Paragon [20] classifies and schedules the application on the proper hardware platform in a way that minimizes interference rather than scheduling the applications in a VM.

Altino Sampaio et al. have proposed new algorithms to dynamically schedule VMs to minimize the performance interference due to hardware resources such as last-level cache(LLC) sharing [19]. In this approach, the aim is to maximize the rate of completed tasks by constructing performance-efficient computing environments that react to performance degradation arise from sharing of LLC memory. And Q-clouds [18], a QoS-aware control framework developed to mitigate performance interference effects. It uses a multi-input multi-output (MIMO) model that tunes resource allocations to capture the performance interference in a virtualized environment. But, Q-Clouds focuses only on the CPU bound workload. In contrast to the above studies [19], [18], though the cloud applications are data centric, it is essential to address the challenges of the I/O interference when running data-intensive applications in virtualized environments. Thus our work focuses on data-intensive applications.

TRACON framework is proposed by Chiang et al.. [1] with the aim of mitigating the interference effects from co-located data-intensive applications, and thus improving the overall system performance. It is composed of an interference prediction component, an interference-aware scheduler, and task and resource monitors. K-means++ algorithm is implemented in interference-aware scheduler to schedule the task in suitable VMs in a virtualized environment. According to Ron C. Chiang et al. in [1], K-means algorithm neither guarantees to converge to a global minimum nor to achieve the best optimal solution available. Genetic Algorithm (GA) is used for the purpose of finding the global minima [14]. Ron C. Chiang et al. [1] proposed the K-means++ algorithm to make a good choice of initial k centers to replace K-means algorithm that is designed to pick points that are far away from each other. Instead of choosing the point farthest from chosen

points, k-means++ pick each point at random with probability proportional to the squared distance. Thus k-means++ is combined with genetic algorithm to find optimized solution. Our proposed work extends this phenomenon by proposing a fast-genetic k-means++ algorithm which is a combination of genetic algorithm with k-means++ to make the optimized decisions in order to improve the overall performance.

### III. PROPOSED WORK

In the proposed work, fast genetic k-means++ is used as a scheduling algorithm to improve the application performance in a virtualized environment. Levenberg-Marquardt method [17], a non-linear model is used to find the optimal solution in predicting the performance. Fast genetic k-means++, a scheduling algorithm is implemented to measure the performance in terms of application throughput, runtime and cost and the progress in the performance results are shown in section 4.

#### A. Terms Used

1) *Interference Prediction Model*: The interference prediction model infers the application performance from the resource consumption observed from multiple VMs. In the proposed work, interference prediction model is constructed using five parameters (controllers) including CPU utilization in VMM, CPU consumption from data processing of application, I/O request, cost and job /cloudlet. These parameters are used to read and write throughput, to measure I/O workload from target application in terms of number of requests per second in order to model the CPU consumption from data processing of the application. The parameter CPU utilization is used to increase the accuracy for a virtualized environment.

a) *Non-linear Model*: According to Ron C. Chiang et al. [1], the prediction accuracy in linear model is mostly at par with weighted mean method, thus it cannot be taken as best fit for the observed data. It is essential to opt for an alternative to linear model and weighted mean method leads us to explore the nonlinear models, in particular with the degree of two that is, quadratic models.

The non-linear model is constructed using Levenberg-Marquardt (LM) Method [17]. It is a combination of two minimizing techniques namely gradient descent method and Gauss-Newton method. The LM method adaptively varies the parameter updates between the gradient descent update and the Gauss-Newton update,

$$[J^T W] + \lambda I] h_{lm} = J^T W y - J$$
 (1)

From the equation, if the parameter  $\lambda$  assumes a small value, then it results in a Gauss-Newton update and if a large value then it takes gradient descent update.

After the above update of the relationship, LM algorithm becomes

$$[J^T W] + \lambda \log(J^T W)] h_{lm} = J^T W y - J$$
 (2)

This algorithm is used to update the parameters in order to obtain optimal solution in virtualized environment. This provides best prediction accuracy for the nonlinear models

$$R = c + \sum_{i=1}^N \alpha_i^{(2)} \cdot P_{VM1,t} + \sum_{i=1}^N \alpha_i^{(3)} \cdot P_{VM2,t} + \sum_{i=1}^N \sum_{j=1}^N \beta_{ij}^{(2)} \cdot P_{VM1,t} \cdot P_{VM2,t} + \sum_{i=1}^N \sum_{j=1}^{i-1} \beta_{ij}^{(3)} \cdot P_{VM1,t} \cdot P_{VM2,t} + \sum_{i=1}^N \sum_{j=1}^N \beta_{ij}^{(4)} \cdot P_{VM1,t} \cdot P_{VM2,t} + \sum_{i=1}^N \sum_{j=1}^{i-1} \beta_{ij}^{(5)} \cdot P_{VM1,t} \cdot P_{VM2,t} + \sum_{i=1}^N \alpha_i^{(4)} \cdot P_{VM1,t}^2 + \sum_{i=1}^N \alpha_i^{(5)} \cdot P_{VM2,t}^2$$
 (3)

The above equation is given for two VMs VM1 and VM2, each one of them can be assigned with one application. Each model of the proposed system architecture relates to five key parameters for individual VMs, thus resulting in ten variables in both VMs together. N is the response variable representing the run time and  $P_{VM1,t}, P_{VM2,t} \in \{1, 2, 3, 4, 5\}$  are the controlled variables representing the application characteristics on VM1 and VM2.

2) *Model training and learning* : Interference profile is generated by running the given application on one VM while the remaining VMs will be executing various workloads in the background where n VMs are involved. This profile has a collection of data on interference effects under different background workloads. This approach supports online learning of interference prediction model that is dynamically modified and monitored for different applications in the cloud platform.

3) *Interference-Aware Scheduling(IAS)*: IAS is proposed for scheduling the task to various VMs with minimized interference effects from co-located applications. It aims to reduce the runtime and improve the I/O throughput for data-intensive applications in a virtualized environment. In the proposed work, fast genetic k-means++ algorithm is used for the purpose of improving the performance in the cloud environment. When the task arrives, the scheduler proceeds with generating number of possible assignments of the same based on the incoming tasks and list of available VMs. Then the scheduling process takes place by assigning the task to different servers based on the predictions.

#### B. Fast genetic k-means ++algorithm (FGKA++)

Fig.1 (a) shows the flow of FGKA++ algorithm which starts with the initialization phase, generating the initial task  $P_0$ . The task in the next generation  $P_{i+1}$  is obtained by applying the following genetic operators sequentially: the selection, the

mutation and the K-means++ on the current task  $P_i$ . The evolution takes place until the termination condition is successfully reached. The algorithm in pseudo code representation is shown in Fig.1 (b)

these approaches could be used to find a globally optimal solution. Genetic Algorithm (GA) is used for the purpose of finding the global minima [13].

K-means algorithm is designed to pick points that are far away from each other. To overcome this drawback, Ron C. Chiang et al. [1] proposed the K-means++ algorithm to make a good choice of initial k centers. Instead of choosing the point farthest from chosen points, k-means++ pick each point at random with probability proportional to the squared distance.

Thus k-means++ is combined with genetic algorithm to find optimized solution. The genetic operators used in this approach are the selection, the distance based mutation and the k-means++ operator are explained below.

1) *The Selection Operator:* The task of the next generation is determined by P independent random processes. Each process randomly selects a solution from the current task  $\{S_1, S_2, \dots, S_P\}$  according to the probability distribution,  $\{R_1, R_2, \dots, R_P\}$  defined by

$$R_p = \frac{F(S_p)}{\sum_{p=1}^P F(S_p)} \quad (p = 1, \dots, P) \quad (4)$$

where  $F(S_p)$  denotes the fitness value of solution  $S_p$ .

2) *The Mutation Operator:* The mutation operator performs the function of shaking the algorithm out of a local optimum, and moving it towards the global optimum. During the mutation,  $b_n^k$  is replaced by  $b_n^{k'}$  for  $n = 1, \dots, N$  simultaneously.  $b_n^{k'}$  is a cluster number randomly selected from  $\{1, \dots, K\}$  with the probability distribution  $\{R_1, R_2, \dots, R_K\}$  defined by

$$R^k = \frac{1.5 * d_{max}(X_n) - d(X_n, c_k) + 0.5}{\sum_{k=1}^K (1.5 * d_{max}(X_n) - d(X_n, c_k) + 0.5)} \quad (5)$$

where  $d(X_n, c_k)$  is the Euclidean distance between pattern  $X_n$  and the centroid  $c_k$  of the  $k^{th}$  cluster, and  $d_{max}(X_n) = \max_k \{d(X_n, c_k)\}$

If the  $k^{th}$  cluster is empty, then  $d(X_n, c_k)$  is defined as 0. The bias 0.5 is introduced to avoid divided-by-zero error in the case that all patterns are equal and are assigned to the same cluster in the given solution.

Initially, the above mutation operator ensures that an arbitrary solution, including the global optimum, might be generated by the mutation from the current solution with a positive probability. Second, it encourages that each  $X_n$  is moving towards a closer cluster with a higher probability. Third, it promotes the probability of converting an illegal solution to a legal one.

3) *K-Means++ Operator:* In order to speed up the convergence process, one step of the classical K-means++ algorithm, which we call K-means++ operator is introduced. Given a solution that is encoded by  $b_1 \dots b_N$ , we replace  $b_n$  by  $b_n'$  for  $n=1, \dots, N$  simultaneously, where  $b_n'$  is the number of the cluster whose centroid is closest to  $X_n$  in Euclidean distance.

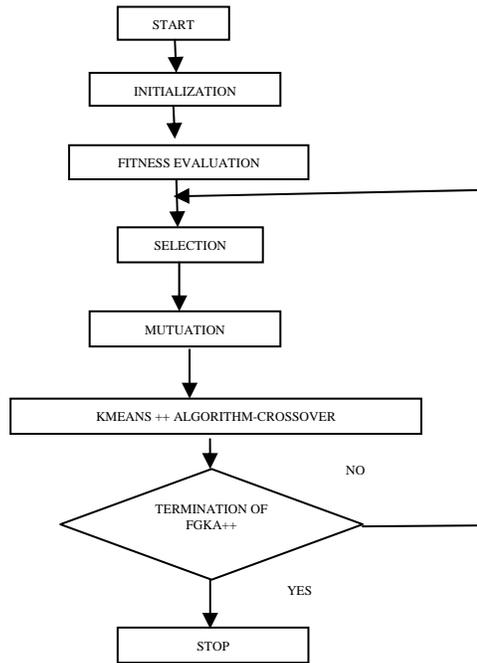


Fig. 1 (a) Flow-chart of FGKA++ algorithm

```

Begin
1. t=0
2. Initialize population P(t)
3. Compute fitness P(t)
4. t=t+1
5. If termination criterion is achieved go to 10
6. Select P(t) from P(t-1)
7. Mutate P(t)
8. K-means++ for crossover P(t)
9. Go to Step 3
10. Output best and stop
End
    
```

Fig. 1 (b) Pseudo code of FGKA++ algorithm

K-Means algorithm (KMA) provides a method of cluster analysis which aims at portioning of n observations into k clusters. Each of the observation belongs to a cluster with the minimum distance between cluster centre and the observation point. It is done iteratively so that the observation point is at least distance from the centre of cluster. The mean distance between the cluster centre and observation is minimized during this iteration process.

The main limitation with the KMA is that, it neither guarantees to converge to a global minimum nor to achieve the best optimal solution available. Since stochastic optimization approaches are good at avoiding convergence to a local optima,

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

In the proposed work, performance is evaluated for own cloud application. The proposed algorithm is compared with K-means ++ algorithm [1] by measuring the performance based on cost, throughput and execution time. In order to generate a more realistic workload, we randomly choose the datasets, data sizes, and number of processes. The comparison result shows the improvements in the performance of virtualized environment.

A. Experimental Setup

Fast genetic k-means++ algorithm is executed on different file types such as pdf, image and text files. The computer hardware used for the implementation is of Intel Core2 duo CPU with 3.40 GHz speed, 4GB RAM size and 500 GB hard disk capacity. The software tools consist of Java as programming language in simulator, Mysql as database and Cloudsim for simulating the virtualized environment.

B. Performance Evaluation

Fast genetic k-means++, scheduling algorithm implemented for scheduling and assigning the tasks to appropriate VMs. When the task arrives, the scheduler proceeds with generating number of possible assignments of the same based on the incoming tasks and list of available VMs. Then the scheduler decides the scheduling and assigning the task to different servers based on the predictions. Thus, the assignment of tasks to suitable VMs with minimum CPU utilization time is experimented on fast genetic k-means++ and k-means++ algorithm and shown the results in Table 1.

Table 1. Scheduling tasks to VMs with CPU utilization on different scheduling algorithms

Fast Genetic k-means++ algorithm			K-means++ algorithm		
Job Name	CPU Utilization	VM Id	Job Name	CPU Utilization	VM Id
Bp.pdf	401.43	4	Bp.pdf	5000	2
Phr2.txt	51.50	1	Phr2.txt	416.67	5
tt.jpg	54.96	6	tt.jpg	250	1

The application throughput is defined as the number of tasks completed in a given period of time. The normalized application throughput is measured for scheduling methods. Fast genetic k-means++ algorithm achieves better throughput compared with k-means++ algorithm is shown in Table 2.

Table 2: Throughput Comparison

Technique	K-means ++	Fast Genetic K-means ++	Improved Throughput
Achieved Throughput (%)	0.62	1.41	0.86

The Throughput, running cost and the execution time taken by the cloud application are measured and the improvement in the performance is shown in Fig.2 &3.

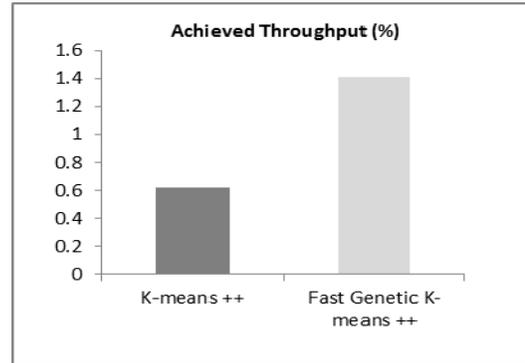


Fig. 2: Throughput Comparison

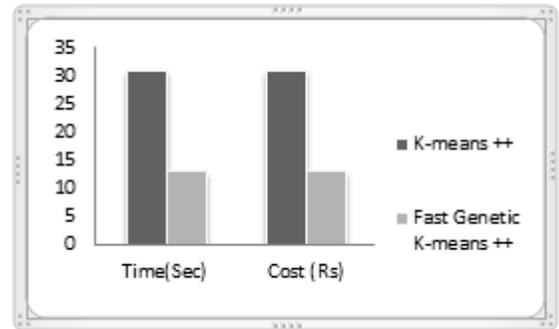


Fig. 3: Comparison of Time and Cost

The cost required for the cloud application is measured based on time utilized in the cloud. The cost of the cloud can be controlled by performing the function with minimal running time. The cost is calculated as. INR 1 per second based on the usage. Hence, the cost for k-means++ is  $31 \times 1 = \text{INR } 31$  whereas, the cost for the fast genetic k-means++ is  $13 \times 1 = \text{INR } 13$ . Thus the proposed technique takes lesser time and lower cost while compared with existing technique and the overall cost depends on the cloud providers. Fast genetic k-means++ technique is compared with k-means++ for the same cloud data and the results are shown in Table 3.

Table 3: Comparison of Time and Cost

Technique	Time (Sec)	Cost (Rs)
K-means ++	31	31.00
Fast Genetic K-means ++	13	13.00

## V. CONCLUSION

The proposed architecture shows effective improvements in terms of throughput, time and cost measures. It is believed that the proposed architecture would highly benefit architectures that use simultaneous running VMs without much performance interferences from others. Our future work involves applying the architecture with new age file types and exploring different scheduling algorithms to further improve the performance in the virtualized environment.

## REFERENCES

- [1]. Ron C. Chiang and H. Howie Huang, "TRACON: Interference-Aware Scheduling for Data-Intensive Applications in Virtualized Environments", IEEE Transactions On Parallel And Distributed Systems, IEEE 2013.
- [2]. Mr. Ajey Singh, Dr. Maneesh Shrivastava, "Overview of Security issues in Cloud Computing", International Journal of Advanced Computer Research (IJACR) Volume 2, Number 1, March 2012.
- [3]. Y. Koh, R. Knauerhase, P. Brett, M. Bowman, Z. Wen, and C. Pu. An Analysis of Performance Interference Effects in Virtual Environments. IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), 2007, pp. 200-209.
- [4]. Ron C. Chiang and H. Howie Huang. TRACON: Interference-aware scheduling for data-intensive applications in virtualized environments. In Proc. Of SC, pages 1–12, nov. 2011.
- [5]. X. Pu, L. Liu, Y. Mei, S. Sivathanu, Y. Koh, and C. Pu. Understanding performance interference of I/O workload in virtualized cloud environments. In Proc. of CLOUD, pages 51–58, July 2010.
- [6]. A. Gulati, G. Shanmuganathan, I. Ahmad, C. Waldspurger, and M. Uysal. Pesto: online storage performance management in virtualized datacenters. In Proc. of SOCC, pages 19:1–19:14, New York, NY, USA, 2011. ACM.
- [7]. H. Shan, K. Antypas, and J. Shalf. Characterizing and predicting the I/O performance of HPC applications using a parameterized synthetic benchmark. In Proc. of SC, pages 42:1–42:12, Piscataway, NJ, USA, 2008. IEEE Press.
- [8]. P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, "Xen and the art of virtualization", in Proceedings of the nineteenth ACM symposium on Operating systems principles, pp. 164-177, 2003.
- [9]. C.A. Waldspurger, "Memory resource management in VMware ESX server", ACM SIGOPS Operating Systems Review, Vol. 36, No. si, p. 181, 2002.
- [10] B.desLigneris, "Virtualization of Linux based computers: the Linux-VServer project", in 19th International Symposium on High Performance

Computing Systems and Applications, pp. 340-346, 2005

- [11] S. J. Vaughan-Nichols, "New approach to virtualization is a lightweight", Computer, Vol. 39, No. 11, pp. 12-14, 2006.
- [12] D. Arthur and S. Vassilvitskii. k-means++: the advantages of careful seeding. In SODA'07.
- [13] K. Krishna and M. NarasimhaMurthy, "Genetic K-Means Algorithm", IEEE Transactions On Systems, Man And Cybernetics—Part B: Cybernetics, Vol. 29, No. 3, June 1999.
- [14] Yi Lu, Shiyong Lu, Farshad Fotouhi, Youping Deng, Susan J. Brown, "FGKA: A Fast Genetic K-means Clustering Algorithm", ACM, 2004.
- [15] Rajkumar Buyya, Rajiv Ranjan, Rodrigo N. Calheiros, "Modeling and Simulation of Scalable Cloud Computing Environments and the CloudSim Toolkit: Challenges And Opportunities", in The International Conference on High Performance Computing and Simulation, HPCS2009, pp:1-11.
- [16] Ravi Iyer, Ramesh Illikkal, Omesh Tickoo, Li Zhao, Padma Apparao, Don Newell, VM3: Measuring, modeling and managing VM shared resources, 2009 Elsevier Computer Networks 53 (2009) 2873–2887.
- [17] K. Madsen, H.B. Nielsen, O. Tingleff. Methods for Non-linear Least Squares problems. Informatics and Mathematical Modelling, Technical University of Denmark. April 2004
- [18] R. Nathuji, A. Kansal and A. Ghaffarkhah. Q-Clouds : managing performance interference effects for qos-aware clouds. In EuroSys '10.
- [19] D. Novaković, Nedeljko Vasić, Stanko Novaković, Dejan Kostić, and Ricardo Bianchini. DeepDive: Transparently Identifying and Managing Performance Interference in Virtualized Environments. Technical Report 183449, EPFL, 2013
- [20] C. Delimitrou et al... Paragon: QoS-aware scheduling for heterogeneous datacenters. In ASPLOS, 2013.
- [21] Altino Sampaio, Jorge G. Barbosa, Parallel & Cloud computing, PCC Vol. 2 Iss. 4, 2013PP. 116-125 www.vkingpub.com © 2013 American V-King Scientific Publishing.

## AUTHORS PROFILE



**A. P. Nirmala** received her MCA and M. Phil., degree in Computer Science from Bharathiar University, Coimbatore in 2000 and 2006 respectively. She is pursuing her Ph.D at Karpagam University, Coimbatore, India. She is presently working as Assistant Professor in the Department of Computer Applications, New Horizon College of Engineering, Bangalore, India. She has 12 years of teaching experience. Her research area is Cloud Computing. She has published 1 research paper in an International Journal and presented 6 research papers in National Conferences.



**Dr. R. Sridaran** is a Ph.D in Computer Science from Madurai Kamaraj University. He has published 25+ research papers in leading journals and presented in many conferences. He is presently guiding eight research scholars in the areas of Cloud Computing, e-learning and Software Engineering. He has got 20 years of academic experience and served in leading educational institutions at different capacities. He is currently the Dean of Faculty of Computer Applications, Marwadi Education Foundation, Rajkot. He is also the Chairman, Computer Society of India, Rajkot Chapter.

# The Comprehensive Review On JDL Model In Data Fusion Networks: Techniques and Methods

Ehsan Azimirad

<sup>1</sup>PHD Student, Electrical and Computer Engineering  
Department, Hakim Sabzevari University  
Sabzevar, Iran

Javad Haddadnia

<sup>2</sup>Associate Professor, Electrical and Computer  
Engineering Department, Hakim Sabzevari University  
Sabzevar, Iran

**Abstract**—this paper provides the comprehensive review on JDL model in multi- sensor data fusion networks and its techniques and methods. Data fusion methods vary greatly depending on the type of problem and the surface to be integrated data. Two main motivations exist for using multiple sensors and combine the results of them. 1) Reduce errors and uncertainty in the measurements, 2) The use of multiple sensors to achieve a better estimate. With the data fusion of multiple databases, or multiple sensors and multiple natures, reduced uncertainty or ambiguity, and reduce complexities. The variety of models and architectures has been provided by researchers to combine sensor data for military and civilian applications. In this paper, we first defined the concept model, architecture and framework based on JDL model, and then states its techniques and methods.

**Keywords**-component; data fusion techniques; JDL model; data fusion models; data fusion architectures

## I. INTRODUCTION

The main task of a battle management system, is integrating a floating sensor output data, target detection, diagnosis and management of a variety of weapons, targets and ultimately the decision is automatically installed on the vessel.

The heart of the battle management systems is data fusion system. Where the type and number of sensors and their reliability is very diverse and the sensors with heterogeneous output cannot overlap nor cannot independent or dependent. For increase the reliability of each of the parts of a battle management system use the fusion algorithms.

Data fusion is the software sector of battle management system. Data fusion is the process of combing information from a number of different sources to provide a robust and complete description of an environment or process of interest. Data fusion is of special significance in any application where a large amount of data must be combined, fused and distilled to obtain information of appropriate quality and integrity on which decisions can be made. Data fusion finds application in many military systems, in civilian surveillance and monitoring tasks, in process control and in information systems.

Data fusion methods are particularly important in the drive toward autonomous systems in all these applications. In principle, automated data fusion processes allow essential measurements and information to be combined to provide knowledge of sufficient richness and integrity that decisions may be formulated and executed autonomously [27].

Data fusion is often (somewhat arbitrarily) divided into a hierarchy of four processes.

Levels 1 and 2 of this process are concerned with the formation of track, identity, or estimate information and the fusion of this information from several sources. Level 1 and 2 fusion is thus generally concerned with numerical information and numerical fusion methods (such as probability theory or kalman filtering).

Level 3 and 4 of the data fusion process is concerned with the extraction of “knowledge” or decisional information. Very often this includes qualitative reporting or secondary sources of information or knowledge from human operators or other sources. Level 3 and 4 fusion is thus concerned with the extraction of high-level knowledge (situation awareness for example) from low level fusion processes, the incorporation of human judgment and the formulation of decisions and actions.

This hierarchy is not, by any means, the only way of considering the general data fusion problem. It is perhaps appropriate for many military data fusion scenarios, but is singularly inappropriate for many autonomous systems or information fusion problems.

The imposition of a “hierarchical” structure to the problem at the outset can also serve to mislead the study of distributed, decentralized and network-centric data fusion structures. Nevertheless, the separate identification of numerical problems (tracking, identification and estimation) from decisional and qualitative problems (situation awareness, qualitative reporting and threat assessment) is of practical value [27].

This paper is organized as follows. Section II is concentrated for statement the concept of model, architecture and framework. Section III has explained data fusion models. This Section is focused on JDL data fusion model and its techniques and methods. Section IV presents the conclusion.

## II. MODEL, ARCHITECTURE AND FRAMEWORK

The conceptual organization of our collected knowledge regarding data fusion has taken many forms. As a result a potential confusion of terminology may arise. We shall therefore define a few terms to describe the way in which data fusion algorithms may be embedded in the context of a larger system.

Three main organizational paradigms are currently in use for describing data fusion systems. These are:

- Models
- Architectures
- Frameworks

We shall describe each of these in turn, highlighting the main differences between them [5]:

**Model** - we define a model, or more specifically a process model, to be a description of a set of processes. This set of processes should be undertaken before the system may be regarded as fully operational. As such it highlights the component functions which the system has but makes no statement regarding their software implementation or physical instantiation.

**Architecture** - we define architecture to be the physical structure of the system. We make particular reference to the way in which data or information is communicated. Architecture includes the arrangement of the component parts, their connectivity and the data flows between them. The architectural description may be high level - data fusion systems which are described as centralized, hierarchical or distributed are classified by their architecture. It may also be specific - blackboard systems [25] and common object request brokering (CORBA) [26] are specific examples of distributed architectures.

**Framework** - we define a framework to be a set of axioms and a reasoning system for manipulating entities based on those axioms. As such a framework provides us with a method of inference from a data-rich / information-poor source to produce abstract concepts which are information-rich. Examples of frameworks currently used in data fusion are probabilistic and evidential reasoning.

The remainder of the paper will concentrate on JDL data fusion model. Architectures and frameworks are equally important but are left for future discussions [5].

## III. MULTISENSOR DATA FUSION MODELS

Data fusion has its roots in the defense research community of the early 1980's. As a result the first data fusion models were either adapted from existing military oriented process models or were designed with a distinctly military flavor [4]. More recently the use of data fusion has broadened to include industrial, medical and commercial applications. More recent models have acknowledged this migration by reducing the military terminology. However, this still exists to some extent (and needs to be changed).

Sensor network configuration, the display information and feedback within the network integration, some of the

major issues in the implementation of a process model are considered.

In Figure 1, a data fusion model is presented for use in various applications. The model in this paper is useful JDL data fusion systems are usually discussed in the context of the military.

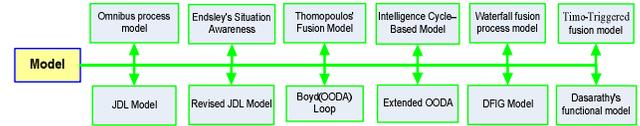


Figure 1. Multi sensor Data Fusion Models

### A. The JDL Model

In the JDL model, proposed by the US Joint Directors of Laboratories Data Fusion Sub-Group in 1985 [28] and updated in 1998 [23], the processing is divided into five levels as depicted in Figure 2.

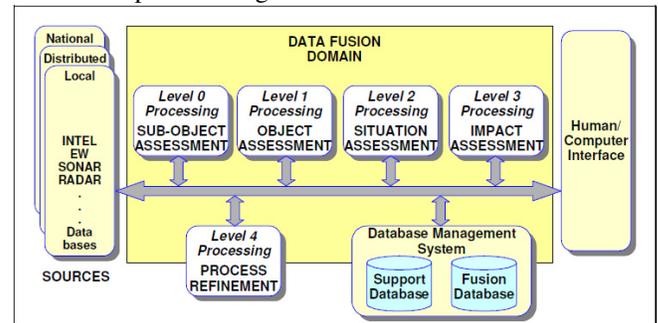


Figure 2. The JDL data fusion model in 1998 [23]

**Level 0** - sub-object data assessment, is associated with pre-detection activities such as pixel or signal processing, spatial or temporal registration.

**Level 1** - object refinement is concerned with the estimation and prediction of continuous (e.g. location or kinematic) or discrete (e.g. behavior or identity) states of objects.

**Level 2** - situation refinement introduces context by examining the relations among entities such as force structure and communication roles. By aggregating objects into meta-objects an interpretation may be placed on the situation.

**Level 3** - implication refinement delineates sets of possible courses of action and the effect they would have in the current situation. This level also introduces the concept that the data fusion system may be operating in an adversarial domain.

**Level 4** - process refinement is an element of resource management and used to close the loop by re-tasking resources (e.g. sensors, communications and processing) in order to support the objectives of the mission. This model has been widely used by the US data fusion community and can now be regarded as the de facto standard for defense data fusion systems, at least in the US. Partly because of its popularity it is applied in a variety of ways [29] and not

always used appropriately.

The goal of the JDL data fusion model is to facilitate understanding and communication among acquisition managers, theoreticians, designers, evaluators, and users of data fusion technology to permit cost-effect system design, development, and operation.

The JDL model was never intended to prescribe a strict ordering on the data fusion levels. This was indicated diagrammatically by the use of an information bus rather than a flow structure. Nevertheless, data fusion system designers have consistently assumed this ordering. Clearly there is a need from users to have an ordering whilst the authors of the JDL model rightly defend the need for a model which admits systems of systems with different hierarchies at different levels.

Process model [7] JDL, processes, functions, classes, different techniques, and special techniques are used to introduce the issue of integration. Figure 3 shows the basic model of the JDL.

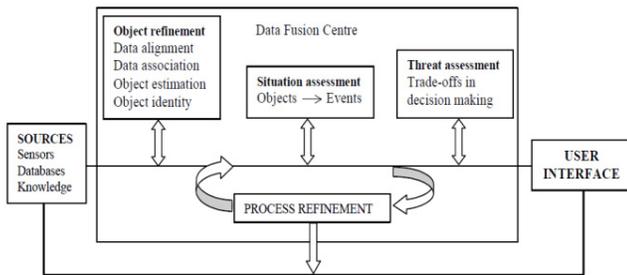


Figure 3. The data fusion framework in JDL model

**Level 0: Sub-Object Assessment:** The task of sub-object data assessment is to reduce the processing load of the fusion processes by pre screening and allocating data to appropriate processes.

**Level 1: Object Assessment:** Level 1 try to combine sensor data to achieve a more accurate and reliable position, velocity, properties, and the identity of the individual objects. This level combines local, parametric, and identity information to obtain representatives of individual objects.

This level performs data alignment (transformation of data to a consistent reference frame and units), association (using correlation methods), tracking actual and future positions of objects, and identification using classification methods.

The overview in data fusion process in this level is in Figure 4.

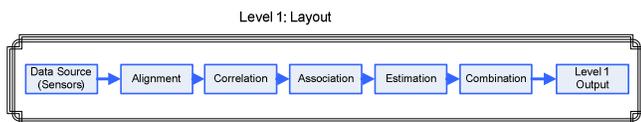


Figure 4. The data fusion process in Level 1

In Figure 5, various techniques have been reviewed to identify targets: statistical pattern recognition methods, the

bayes estimation method, Dempster- Shafer evidence theory, fuzzy integral theory and neural network approach.

In general, there are a variety of methods for data fusion that are classified as follows [21]:

1. The data fusion techniques that reduce data dimension, such as principal component analysis (PCA), sammon mapping, artificial neural networks (ANNs) and data clustering.
2. The direct fusion techniques, such as simple averaging techniques and estimation techniques such as the family of kalman filtering or partial filters and neural networks.
3. The feature fusion techniques such as neural networks, clustering, statistical classifiers, support vector machine (SVM) and methods of mapping.
4. The decision fusion techniques include weighted voting, consensus and etc., methods based on artificial intelligence and inference methods based on probabilities such as Bayesian or based on belief and possibility such as demster- Shafer.

Using of the common algorithms in target identification like Dempster- Shafer and Bayesian is very difficult in problem that sensors used in the system not independent or reliable information fusion is not enough [23]. As a result, these methods produce a large error by the complexity of the system and more sophisticated system or violent battle field. In such circumstances, and especially during the phase of incomplete data, produces better results.

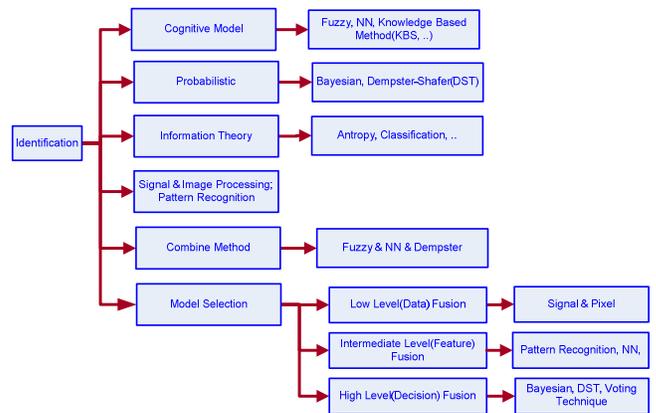


Figure 5. The different methods for target identification in JDL Level 1

Dempster- Shafer theory is for use in decision fusion that, in the 1990s as a safe and reliable method and as a special case of the Bayes filter was proposed. The weakness of this method is that data sources have a huge contradiction to each other and the similarity of the data production is low.

In the problems of multi-sensor estimating, it is required that a large number of measurements assigned to related states. In Figure 6, the various methods of data assignment are listed.

TABLE I. THE TECHNIQUES OF ONE LEVEL OF JDL MODEL

JDL Process	Processing function	Techniques
Level 1: Object Refinement	Data alignment	<ul style="list-style-type: none"> <li>• Coordinate transforms</li> <li>• Units Adjustments</li> </ul>
	Data/object correlation	<ul style="list-style-type: none"> <li>• Gating techniques</li> <li>• Multiple hypothesis association</li> <li>• Probabilistic data association</li> <li>• Nearest neighbor</li> </ul>
	Position/kinematic and attribute estimation	<ul style="list-style-type: none"> <li>• Sequential estimation                             <ul style="list-style-type: none"> <li>- Kalman filter</li> <li>- <math>\alpha\beta</math> filter</li> </ul> </li> <li>• Multiple hypothesis</li> <li>• Batch estimation</li> <li>• Maximum likelihood</li> <li>• Hybrid methods</li> </ul>
	Object identity estimation	<ul style="list-style-type: none"> <li>• Physical models</li> <li>• Feature-based techniques                             <ul style="list-style-type: none"> <li>- Neural networks</li> <li>- Cluster algorithms</li> <li>- Pattern recognition</li> </ul> </li> <li>• Syntactic models</li> </ul>

The challenges and limitations contained in Level 1 are busy goal environments, targets with quickly maneuver, releasing a complex signals, sensor correlated observations, noise and clutter background, multi-streaming environments, environments with interference channel, automatic detection of content data and data aggregation kinetics and identity.

*Level 2: Situation Assessment:* Situation Refinement attempts to develop a description of the relationship between objects and observed events. This level attempts to construct a picture from incomplete information provided by level 1, to relate the reconstructed entity with an observed event.

Situation assessment are used of the knowledge-based algorithms such as rule-based expert systems, fuzzy logic, knowledge-based systems (KBS), neural networks, dynamics Bayesian networks (DBN) and Dempster-Shaffer network (DSN). These algorithms are not yet evolved. The main challenge in this area is to collect a series of knowledge-based framework, documents, and other ways to convey information about the evaluation conditions. Unfortunately, existing inference models to mimic human performance in this area are basic models. More research is needed to obtain reliable knowledge-based systems in large areas, to assess the situation created automatically. The new trends that can be used are fuzzy logic and hybrid structures. Some of the common algorithms in this area are shown in Figure 8.

The challenges and limitations contained in Level 2 are inefficient primitive Prototypes, lack of tests on scale field models, the difficult development of knowledge-based methods and the lack of law mechanism for data fusion testing and evaluation.

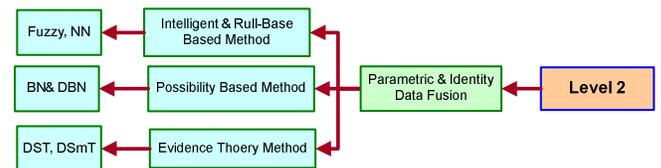


Figure 8. The techniques of JDL Level 2

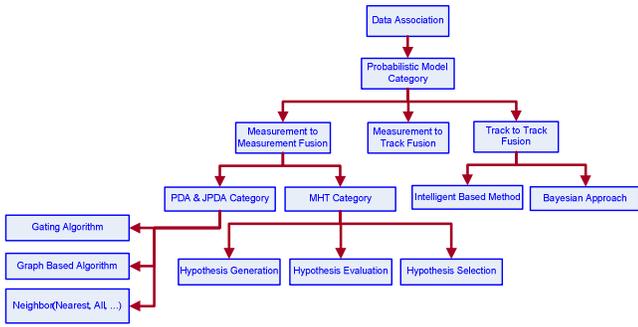


Figure 6. The different methods for target assignment in JDL Level 1

The sensor output in some cases is Measurement and in some cases to the Track. There are two common methods for calculating the data to data acquisition, the data appointment of probabilistic methods (family JPDA) and methods of multi hypotheses (MHT). In MHT method, production, evaluation and selection hypotheses are the most problems, while probabilistic methods are considered gating algorithms and are used on similar approach to their nearest neighbors.

In estimation and target tracking, there are traditional systems based on the use of probability-based methods such as Kalman filter family and particulate filters, while the new systems are used methods of based on finite set theory such as the family of probability density filter (PHD). The advantage of PHD filter compared to traditional forms is better answers and faster convergence toward similar situation. The methods such as Dempster-Shaffer and Dizert-Asmarendakh are used as combining the decision results in tracking part. Some of the conventional methods of estimation and tracking problem solving are given in Figure 7.

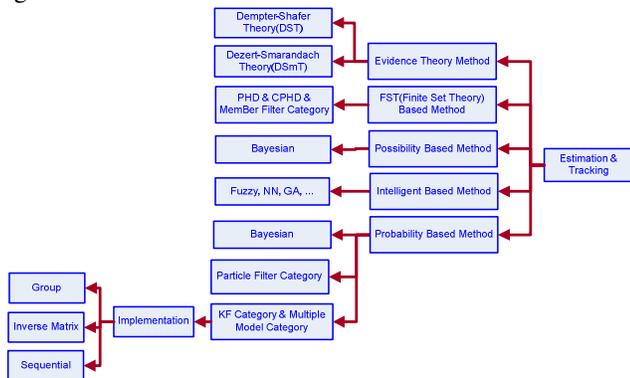


Figure 7. The different methods for target tracking in JDL Level 1

The Most perfect of data fusion process in JDL model is Level 1. This means that this level is using for positioning, velocity, profile, and the object identity. The conventional methods that used in JDL model has mentioned TABLE I.

**Level 3: Threat Assessment:** Threat assessment can be one of the most crucial stages of the battle management system in data fusion. An incorrect threat assessment can cost the loss of some features or even threaten the life. A large number of articles and research from 2001 onwards was devoted to it. It would coincide with the relative maturity of the methods and data fusion algorithms at low levels, namely the situation assessment [9-13].

Threat Refinement Based on a priori knowledge and predictions about the future situation this processing level, tries to draw inferences about vulnerabilities and opportunities for operation. This level interprets the results from level 2 in terms of the possible opportunities for operation. It analyses the advantages and disadvantages of taking one course of action over another. The threat assessment consists of several consecutive steps:

*a) Threat Estimation*

Threat estimation monitor target threat value from three aspects: time and opportunity to destroy by enemy and the amount of space and target near time to hypothetical situation from a certain perspective for example itself floating [14-17].

This step requires data such as speed, distance, closest point of approach (CPA) to enemy and time to reach it. A variety of methods have been developed by researchers that are weighted the data. The overall goal, to derive the threats that may be misleading, for example, targets to their own direction is not to itself float. Targets are not offense intended goals or are not considered due to their low speed.

*b) Ability to enemy destroys, and analyzes its weapons*

To accomplish this task, a very complete database is required to calculate and evaluate some of the steps in the following possible actions:

1. The similarity of the same threats, using an algorithm such as Dempster- Shaffer is achievable. Characteristics and abilities identified targets linked to the richness of the database.
2. Calculate the arrival time of the most threatened (high range) of the enemy's weapons to itself vessel, assuming a constant velocity target.
3. Calculate the time to reach the most threatened (high range) floating weapons

To answer the above cases, the following data are needed: the similarity of threats, the similarity of weapons that are more threaten, weapon speed, maximum range of the weapon, weapons type (missiles, torpedoes, or otherwise), the ability of weapons (anti-surface or under-surface or air).

*c) Target Intent*

Intent assessment is the most challenging issues in this field. The models and scenarios for the assessment of enemy and friendly units responded to the doctrine of probable enemy in war, one of the new issues and developing that it demands programming algorithms such as game theory. Also, the

difficult deduce of conditions with rapid change is one of the appropriate fields in this area.

*d) The involvement Design*

Finally, after the previous steps and gather the necessary information review the conflict design with the enemy ships and weapons assignment to them with the priority level of each target threat to own float.

There are varieties of methods for threat assessment in data fusion. They are such as neural networks, fuzzy logic and Bayesian belief networks. Some of the most widely methods used in this area is shown in Figure 9.

The challenges and limitations contained in Level 3 are similar to level 2 including of inefficient primitive Prototypes, lack of tests on scale field models, the difficult development of knowledge-based methods and the lack of law mechanism for data fusion testing and evaluation.

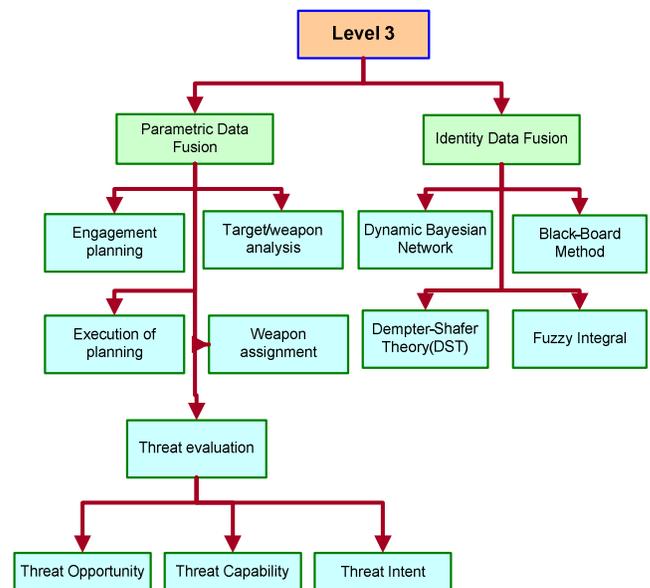


Figure 9. The techniques of JDL Level 3

TABLE II shows the functions, processes and necessary techniques to conduct a threat assessment in the third level of JDL model.

TABLE II. THE PROCESSING FUNCTIONS AND THREAT ASSESSMENT TECHNIQUES IN JDL MODEL

JDL Process	Processing Function	Techniques
Level 3: Threat Refinement	Aggregate force estimation	Neural networks
	Intent prediction	Blackboard systems
	Multi- perspective assessment	Fast- time engagement models

Figure 10 depicts the different entities associated with threat assessment and the relationships between those entities. The proposed taxonomy consists of a high-level

threat that is composed of one or more threat signatures.

A threat signature is an attribute or property of the threat that is detectable. A threat observable is an embodiment of the threat signature that reveals the presence of the signature. While the terms signature and observable are more typically associated with the physical characteristics of an object, we extend this language to include nonphysical observables as well. In threat assessment, the detection of an observable such as the frequency of a term/phrase in a collection of documents may be as significant as detecting the density of a material.

Both the physical and non-physical signatures of threats are treated equally in the threat taxonomy. Its structure does not attempt to convey significance of each threat signature, as that is handled through the Threat Assessment Engine [22].

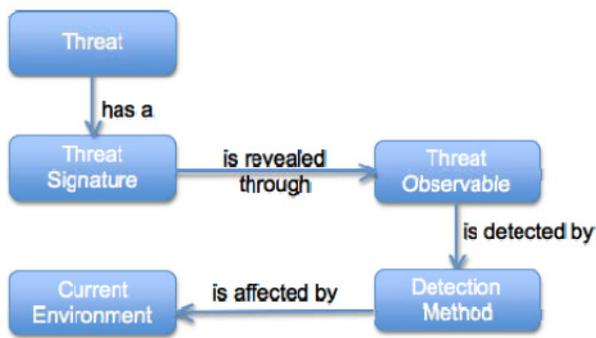


Figure 10. Threat Classification

**Level 4: Process Refinement:** Level 4 is call processing to monitor the whole process of data fusion, to evaluate and improve the performance of real-time systems. Process Refinement is a process concerned with other processes. The three key functions are:

- Monitor the real-time and long-term data fusion performance.
- Identify information required to improve the multi-level data fusion.
- Allocate and direct sensors and sources to achieve the mission goals.

The techniques JDL models in level 4 are shown in figure 11.

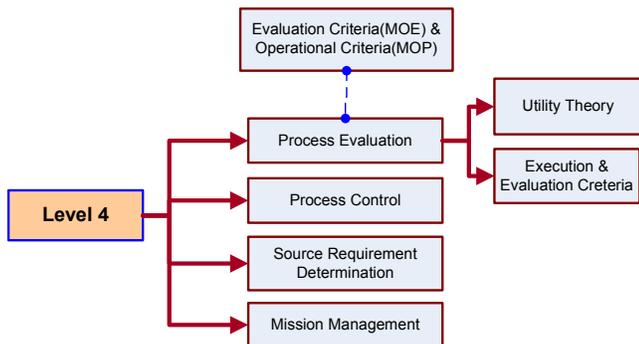


Figure 11. The techniques of JDL Level 4

TABLE III shows briefly the conventional methods in two, three and four level of JDL data fusion model.

TABLE III. THE TECHNIQUES OF 2, 3 AND 4 LEVEL OF JDL MODEL

Level 2:Situation Refinement	Object aggregation Event/activity interpretation Contextual interpretation	<ul style="list-style-type: none"> <li>• Knowledge-based systems(KBS)</li> <li>- Rule-based expert systems</li> <li>- Fuzzy logic</li> <li>- Frame-based</li> <li>• Logical templating</li> <li>• Neural networks</li> <li>- Blackboard systems</li> </ul>
Level 3:Threat Refinement	Aggregate force estimation Intent prediction Multi-perspective assessment	<ul style="list-style-type: none"> <li>• Neural networks</li> <li>- Blackboard systems</li> <li>• Fast-time engagement models</li> </ul>
Level 4:Process Refinement	Performance evaluation	<ul style="list-style-type: none"> <li>• Measure of evaluation</li> <li>• Measures of performance</li> <li>• Utility theory</li> </ul>
	Process control	<ul style="list-style-type: none"> <li>• Multi-objective optimization</li> <li>- Linear programming</li> <li>- Goal programming</li> </ul>
	Source requirement determination	<ul style="list-style-type: none"> <li>• Sensor models</li> </ul>
	Mission management	<ul style="list-style-type: none"> <li>• Knowledge-based systems</li> </ul>

**Data Management System:** this system is for storage and retrieval of pre-processed data and human-computer interaction. Its task is to monitor, evaluate, add, update and provide information for the fusion processes.

**Human-computer Interaction:** This part provides an interface for human input and communication of fusion results to operators and users. Some of the problems of JDL model include:

- Lack of attention to the problems of combining multiple images.
- Lack of support for multi-section sensors.

In this model, zero-level was interpreted as the features object assessment level that estimation and prediction signal or object is done it.

In 2003, Mr. Blasch and Plano presented DFIG model for the development of JDL model. Accordingly, level 5 name user refinement that is related to a combination of human interface and process control information is added to the JDL model levels. So, the levels of 0 to 3 are the previous four levels and the levels of 4 and 5 are developed as the two new surfaces. Level 4 in resource Management and Level 5 in human-machine interface functions are defined. Figure 12 illustrates this model [24].

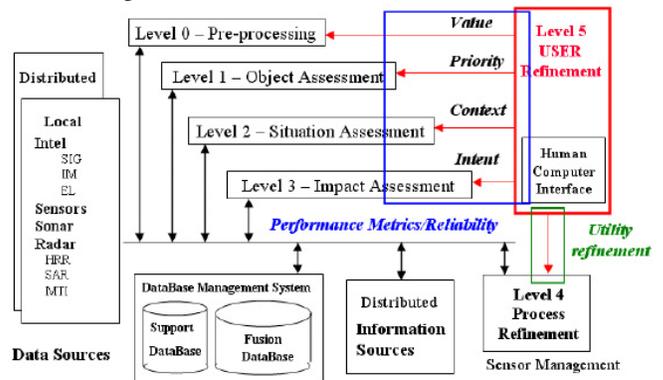


Figure 12. User data fusion model

Based on user fusion model, DFIG model is designed with the aim of separating the between data fusion and

management functions. Management functions are divided into three sections sensor control, placing the platform and user choice for achieving mission objectives. Figure 15 shows the model.

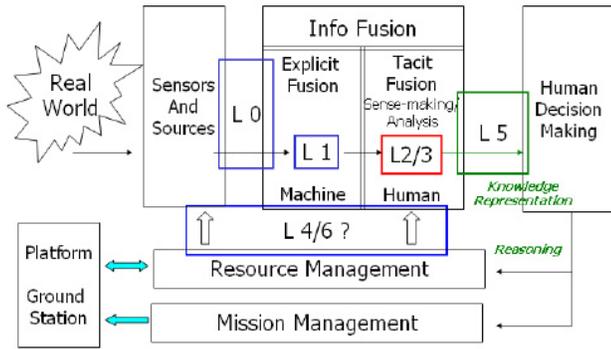


Figure 13. DFIG model

Level 2 includes tacit functions which are inferred from level 1 explicit representations of object assessment. Since the unobserved aspects of the SA problem can not be processed by a computer, user knowledge and reasoning is necessary. The current definitions, based on the revised JDL fusion model [5], include [24]:

- Level 0 – Data Assessment: estimation and prediction of signal/object observable states on the basis of pixel/signal level data association (e.g. information systems collections);
- Level 1 – Object Assessment: estimation and prediction of entity states on the basis of data association, continuous state estimation and discrete state estimation (e.g. data processing);
- Level 2 – Situation Assessment: estimation and prediction of relations among entities, to include force structure and force relations, communications, etc. (e.g. information processing);
- Level 3 – Impact Assessment: estimation and prediction of effects on situations of planned or estimated actions by the participants; to include interactions between action plans of multiple players (e.g. assessing threat actions to planned actions and mission requirements, performance evaluation);
- Level 4 – Process Refinement (an element of Resource Management): adaptive data acquisition and processing to support sensing objectives (e.g. sensor management and information systems dissemination, command / control).
- Level 5 – User Refinement (an element of Knowledge Management): adaptive determination of who queries information and who has access to information (e.g. information operations) and adaptive data retrieved and displayed to support cognitive decision making and actions (e.g. human computer interface).
- Level 6 – Mission Management (an element of Platform Management): adaptive determination of

spatial-temporal control of assets (e.g. airspace operations) and route planning and goal determination to support team decision making and actions (e.g. theater operations) over social, economic, and political constraints.

#### IV. CONCLUSION

The goal of the JDL data fusion model is to facilitate understanding and communication among acquisition managers, theoreticians, designers, evaluators, and users of data fusion technology to permit cost-effect system design, development, and operation. It should be emphasized that the model was conceived as a functional model, not as a process model or as an architectural paradigm.

The JDL model has Generic Architecture. A generic architecture gives an outline how to implement an application, but for example does not specify which operating system, hardware, communication system or database should be used.

Achieve a comprehensive understanding of the methods and algorithms used for data fusion is one of the necessary steps for the practical implementation of this model.

In this paper, JDL data fusion model is fully described and its processing levels were discussed. An overview of the methods and algorithms of processing was done and new models are introduced to correct existing deficiencies in the model.

#### REFERENCES

- [1] J. Esteban, A. Starr, R. Willetts, P. Hannah and P. B. Cross, "A Review of data fusion models and architectures: towards engineering guidelines", p50c 10-November-2004
- [2] S. Tzu, "The Art of War", [Internet] 134-118 BC [cited 2004 Nov 6]; chapter 3, section 18. Available from: <http://classics.mit.edu/Tzu/artwar.html>.
- [3] K. Emery, "Surface Navy Combat Systems Engineering Strategy", Chief Architect PEO Integrated Warfare Systems, 4 March 2010.
- [4] P. Valin, E. Bosse, and A. Jouan. "Airborne application of information fusion algorithms to classification.", Technical Report TR 2004-282, Defense Research and Development Canada – Valcartier, May 2006.
- [5] M. Bedworth and J. OBrien, "The Omnibus Model: A New Model of Data Fusion?", 1999.
- [6] Y. Liang. "An Approximate Reasoning Model for Situation and Threat Assessment." In Proceedings of the 4th International Conference on Fuzzy Systems and Knowledge Discovery, 2007.
- [7] E. Franklin and Jr. White, "Data Fusion Lexicon", Joint Directors of Laboratories, Technical Panel for C3, Data Fusion Sub-Panel, Naval Ocean Systems Center, San Diego, 1987.
- [8] J. Llinas and D. L. Hall, "An introduction to multi-sensor data fusion.", PROCEEDINGS OF THE IEEE, VOL. 85, NO. 1, JANUARY 1997.
- [9] S.J. Yang, J. Holsopple, and M. Sudit. "Evaluating Threat Assessment for Multi-Stage Cyber Attacks." In Proceedings of the 2006 Military Communications Conference. Washington, DC. Oct. 23-25, 2006.
- [10] R. Chinchani, A. Iyer, H.Q. Ngo, and S. Upadhyaya. "Towards a theory of insider threat assessment." In Proceedings of the 2005 International Conference on Dependable Systems and Networks, 2005.

- [11] F. BOLDERHEIJ, PHD thesis, "Mission Driven" Netherlands Defence Academy and the Centre for Automation of Mission Critical Systems (CAMS) – Force Vision. 2007.
- [12] GA. McIntyre and KJ. Hintz, "A Comprehensive Approach to Sensor Management, Part I: A Survey of Modern Sensor Management Systems", IEEE Transactions on SMC, April 1999.
- [13] A. Erhard and S. McGalliard, "ADVANCES IN MILITARY MULTI-SENSOR DATA FUSION TECHNOLOGY AND APPLICATIONS FOR CIVILIAN USE", Eleventh Annual Freshman Conference, April 9, 2011.
- [14] J. Llinas, C. Bowman, G. Rogova, A. Steinberg, E. Waltz and F. White, "Revisiting the JDL Data Fusion Model II", 1999.
- [15] B.V. Dasarathy, "Decision Fusion", IEEE Computer Society Press, 1994.
- [16] Y. Liang. "An Approximate Reasoning Model for Situation and Threat Assessment." In Proceedings of the 4th International Conference on Fuzzy Systems and Knowledge Discovery, 2007.
- [17] R. Chinchani, A. Iyer, H.Q. Ngo, and S. Upadhyaya. "Towards a theory of insider threat assessment." In Proceedings of the 2005 International Conference on Dependable Systems and Networks, 2005.
- [18] S.K. Kashyap and J.R. Raol, "Fuzzy Logic Applications in Filtering and Fusion for Target Tracking", Defence Science Journal, Vol. 58, No. 1, January 2008, pp. 120-135.
- [19] N. Rao, P. Sudesh, K. Kashyap and G. Girija, "SITUATION ASSESSMENT IN AIRCOMBAT: A FUZZY-BAYESIAN HYBRID APPROACH", Proceedings of the International Conference on Aerospace Science and Technology 26-28 June 2008, Bangalore, India.
- [20] C. Lollett, PHD Thesis, "Belief Based Reinforcement Learning For Data Fusion", January, 2009.
- [21] A. N. Steinberg, "Threat Assessment Technology Development", Springer-Verlag Berlin Heidelberg, USA, 2005.
- [22] M. B. Justin, A. K. Ryan and N. T. Jim, "An Information Fusion Framework for Threat Assessment", Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6285.
- [23] A. Steinberg, C. Bowman, and F. White, "Revisions to the JDL Data Fusion Model", SPIE, Vol. 3719, pp. 430-441, 1999.
- [24] E. Blasch and S. Plano, "DFIG Level 5 (User Refinement) issues supporting Situational Assessment Reasoning", Fusion 05, July 2005.
- [25] E. Shahbazian, E. Bosse and P. Valin, "Multi-Agent Data Fusion Workstation (MADFW) Architecture", Proceedings of AeroSense conference, SPIE Vol. 3376, pp. 60-68 (1998).
- [26] M. Balci and S. Kuru, "A CORBA Based Infrastructure (CORBIS) for Sensor Data Fusion Systems", Proceedings of AeroSense conference, SPIE Vol. 3179, pp. 220-229 (1999).
- [27] H. Durrant-Whyte, "Multi Sensor Data Fusion", Australian Centre for Field Robotics The University of Sydney NSW 2006, January 2001.
- [28] F. White, "A Model for Data Fusion", Proceedings 1st National Symposium on Sensor Fusion (1988).
- [29] L. Klein, "Sensor and Data Fusion Concepts and Applications", SPIE Volume TT14 (1993).



**Javad Haddadnia**, received his B.S. and M.S. degrees in electrical and electronic engineering with the first rank from Amirkabir University of Technology, Tehran, Iran, in 1993 and 1995, respectively. He received his Ph.D. degree in electrical engineering from Amirkabir University of Technology, Tehran, Iran in 2002. He joined Hakim Sabzevari University in Iran. His research interests include neural network, fuzzy logic and its applications in data fusion, threat assessment and any other problems, digital image processing, computer vision, and face detection and recognition. He has published several papers in these areas. He has served as a Visiting Research Scholar at the University of Windsor, Canada during 2001-2002. He is a member of SPIE, CIPPR, and IEICE.

#### AUTHORS PROFILE



**Ehsan Azimirad**, received the B.Sc. degree in computer engineering and M.Sc. degree in control engineering with honors from the Ferdowsi University of Mashhad, Mashhad, Iran, in 2006 and 2009, respectively. He is now PHD student in electrical and electronic engineering at Hakim Sabzevari University of Sabzevar in Iran. His research interests are fuzzy control systems and its applications in urban traffic, data fusion, threat assessment and any other problems, nonlinear control, Image Processing and Pattern Recognition and etc.

## IJCSIS AUTHORS' & REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA  
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia  
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA  
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway  
Assoc. Prof. N. Jaisankar, VIT University, Vellore, Tamilnadu, India  
Dr. Amogh Kavimandan, The Mathworks Inc., USA  
Dr. Ramasamy Mariappan, Vinayaka Missions University, India  
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China  
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA  
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico  
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India  
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania  
Dr. Junjie Peng, Shanghai University, P. R. China  
Dr. Ilhem LENGILIZ, HANA Group - CRISTAL Laboratory, Tunisia  
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India  
Dr. Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain  
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India  
Dr Li Fang, Nanyang Technological University, Singapore  
Prof. Pijush Biswas, RCC Institute of Information Technology, India  
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia  
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India  
Dr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand  
Dr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India  
Dr. Hayder N. Jasem, University Putra Malaysia, Malaysia  
Dr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India  
Dr. R. S. Karthik, C. M. S. College of Science and Commerce, India  
Dr. P. Vasant, University Technology Petronas, Malaysia  
Dr. Wong Kok Seng, Soongsil University, Seoul, South Korea  
Dr. Praveen Ranjan Srivastava, BITS PILANI, India  
Dr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong  
Dr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia  
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan  
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria  
Dr. Riktresh Srivastava, Skyline University, UAE  
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia  
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt  
and Department of Computer science, Taif University, Saudi Arabia  
Dr. Tirthankar Gayen, IIT Kharagpur, India  
Dr. Huei-Ru Tseng, National Chiao Tung University, Taiwan

Prof. Ning Xu, Wuhan University of Technology, China  
Dr Mohammed Salem Binwahlan, Hadhramout University of Science and Technology, Yemen  
& Universiti Teknologi Malaysia, Malaysia.  
Dr. Aruna Ranganath, Bhoj Reddy Engineering College for Women, India  
Dr. Hafeezullah Amin, Institute of Information Technology, KUST, Kohat, Pakistan  
Prof. Syed S. Rizvi, University of Bridgeport, USA  
Dr. Shahbaz Pervez Chattha, University of Engineering and Technology Taxila, Pakistan  
Dr. Shishir Kumar, Jaypee University of Information Technology, Wakanaghat (HP), India  
Dr. Shahid Mumtaz, Portugal Telecommunication, Instituto de Telecomunicações (IT) , Aveiro, Portugal  
Dr. Rajesh K Shukla, Corporate Institute of Science & Technology Bhopal M P  
Dr. Poonam Garg, Institute of Management Technology, India  
Dr. S. Mehta, Inha University, Korea  
Dr. Dilip Kumar S.M, University Visvesvaraya College of Engineering (UVCE), Bangalore University,  
Bangalore  
Prof. Malik Sikander Hayat Khiyal, Fatima Jinnah Women University, Rawalpindi, Pakistan  
Dr. Virendra Gomase , Department of Bioinformatics, Padmashree Dr. D.Y. Patil University  
Dr. Irraivan Elamvazuthi, University Technology PETRONAS, Malaysia  
Dr. Saqib Saeed, University of Siegen, Germany  
Dr. Pavan Kumar Gorakavi, IPMA-USA [YC]  
Dr. Ahmed Nabih Zaki Rashed, Menoufia University, Egypt  
Prof. Shishir K. Shandilya, Rukmani Devi Institute of Science & Technology, India  
Dr. J. Komala Lakshmi, SNR Sons College, Computer Science, India  
Dr. Muhammad Sohail, KUST, Pakistan  
Dr. Manjaiah D.H, Mangalore University, India  
Dr. S Santhosh Baboo, D.G.Vaishnav College, Chennai, India  
Prof. Dr. Mokhtar Beldjehem, Sainte-Anne University, Halifax, NS, Canada  
Dr. Deepak Laxmi Narasimha, Faculty of Computer Science and Information Technology, University of  
Malaya, Malaysia  
Prof. Dr. Arunkumar Thangavelu, Vellore Institute Of Technology, India  
Dr. M. Azath, Anna University, India  
Dr. Md. Rabiul Islam, Rajshahi University of Engineering & Technology (RUET), Bangladesh  
Dr. Aos Alaa Zaidan Ansaef, Multimedia University, Malaysia  
Dr Suresh Jain, Professor (on leave), Institute of Engineering & Technology, Devi Ahilya University, Indore  
(MP) India,  
Dr. Mohammed M. Kadhum, Universiti Utara Malaysia  
Dr. Hanumanthappa. J. University of Mysore, India  
Dr. Syed Ishtiaque Ahmed, Bangladesh University of Engineering and Technology (BUET)  
Dr Akinola Solomon Olalekan, University of Ibadan, Ibadan, Nigeria  
Dr. Santosh K. Pandey, Department of Information Technology, The Institute of Chartered Accountants of  
India  
Dr. P. Vasant, Power Control Optimization, Malaysia  
Dr. Petr Ivankov, Automatika - S, Russian Federation

Dr. Utkarsh Seetha, Data Infosys Limited, India  
Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal  
Dr. (Mrs) Padmavathi Ganapathi, Avinashilingam University for Women, Coimbatore  
Assist. Prof. A. Neela madheswari, Anna university, India  
Prof. Ganesan Ramachandra Rao, PSG College of Arts and Science, India  
Mr. Kamanashis Biswas, Daffodil International University, Bangladesh  
Dr. Atul Gonsai, Saurashtra University, Gujarat, India  
Mr. Angkoon Phinyomark, Prince of Songkla University, Thailand  
Mrs. G. Nalini Priya, Anna University, Chennai  
Dr. P. Subashini, Avinashilingam University for Women, India  
Assoc. Prof. Vijay Kumar Chakka, Dhirubhai Ambani IICT, Gandhinagar ,Gujarat  
Mr Jitendra Agrawal, : Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal  
Mr. Vishal Goyal, Department of Computer Science, Punjabi University, India  
Dr. R. Baskaran, Department of Computer Science and Engineering, Anna University, Chennai  
Assist. Prof, Kanwalvir Singh Dhindsa, B.B.S.B.Engg.College, Fatehgarh Sahib (Punjab), India  
Dr. Jamal Ahmad Dargham, School of Engineering and Information Technology, Universiti Malaysia Sabah  
Mr. Nitin Bhatia, DAV College, India  
Dr. Dhavachelvan Ponnurangam, Pondicherry Central University, India  
Dr. Mohd Faizal Abdollah, University of Technical Malaysia, Malaysia  
Assist. Prof. Sonal Chawla, Panjab University, India  
Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India  
Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia  
Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia  
Professor Dr. Sabu M. Thampi, .B.S Institute of Technology for Women, Kerala University, India  
Mr. Noor Muhammed Nayeem, Université Lumière Lyon 2, 69007 Lyon, France  
Dr. Himanshu Aggarwal, Department of Computer Engineering, Punjabi University, India  
Prof R. Naidoo, Dept of Mathematics/Center for Advanced Computer Modelling, Durban University of  
Technology, Durban,South Africa  
Prof. Mydhili K Nair, M S Ramaiah Institute of Technology(M.S.R.I.T), Affiliated to Visweswaraiah  
Technological University, Bangalore, India  
M. Prabu, Adhiyamaan College of Engineering/Anna University, India  
Mr. Swakkhar Shatabda, Department of Computer Science and Engineering, United International University,  
Bangladesh  
Dr. Abdur Rashid Khan, ICIT, Gomal University, Dera Ismail Khan, Pakistan  
Mr. H. Abdul Shabeer, I-Nautix Technologies,Chennai, India  
Dr. M. Aramudhan, Perunthalaivar Kamarajar Institute of Engineering and Technology, India  
Dr. M. P. Thapliyal, Department of Computer Science, HNB Garhwal University (Central University), India  
Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran  
Mr. Zeashan Hameed Khan, : Université de Grenoble, France  
Prof. Anil K Ahlawat, Ajay Kumar Garg Engineering College, Ghaziabad, UP Technical University, Lucknow  
Mr. Longe Olumide Babatope, University Of Ibadan, Nigeria  
Associate Prof. Raman Maini, University College of Engineering, Punjabi University, India

Dr. Maslin Masrom, University Technology Malaysia, Malaysia  
Sudipta Chattopadhyay, Jadavpur University, Kolkata, India  
Dr. Dang Tuan NGUYEN, University of Information Technology, Vietnam National University - Ho Chi Minh City  
Dr. Mary Lourde R., BITS-PILANI Dubai , UAE  
Dr. Abdul Aziz, University of Central Punjab, Pakistan  
Mr. Karan Singh, Gautam Budtha University, India  
Mr. Avinash Pokhriyal, Uttar Pradesh Technical University, Lucknow, India  
Associate Prof Dr Zuraini Ismail, University Technology Malaysia, Malaysia  
Assistant Prof. Yasser M. Alginahi, College of Computer Science and Engineering, Taibah University, Madinah Munawwarrah, KSA  
Mr. Dakshina Ranjan Kisku, West Bengal University of Technology, India  
Mr. Raman Kumar, Dr B R Ambedkar National Institute of Technology, Jalandhar, Punjab, India  
Associate Prof. Samir B. Patel, Institute of Technology, Nirma University, India  
Dr. M.Munir Ahamed Rabbani, B. S. Abdur Rahman University, India  
Asst. Prof. Koushik Majumder, West Bengal University of Technology, India  
Dr. Alex Pappachen James, Queensland Micro-nanotechnology center, Griffith University, Australia  
Assistant Prof. S. Hariharan, B.S. Abdur Rahman University, India  
Asst Prof. Jasmine. K. S, R.V.College of Engineering, India  
Mr Naushad Ali Mamode Khan, Ministry of Education and Human Resources, Mauritius  
Prof. Mahesh Goyani, G H Patel Collge of Engg. & Tech, V.V.N, Anand, Gujarat, India  
Dr. Mana Mohammed, University of Tlemcen, Algeria  
Prof. Jatinder Singh, Universal Institutiion of Engg. & Tech. CHD, India  
Mrs. M. Anandhavalli Gauthaman, Sikkim Manipal Institute of Technology, Majitar, East Sikkim  
Dr. Bin Guo, Institute Telecom SudParis, France  
Mrs. Maleika Mehr Nigar Mohamed Heenaye-Mamode Khan, University of Mauritius  
Prof. Pijush Biswas, RCC Institute of Information Technology, India  
Mr. V. Bala Dhandayuthapani, Mekelle University, Ethiopia  
Dr. Irfan Syamsuddin, State Polytechnic of Ujung Pandang, Indonesia  
Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius  
Mr. Ravi Chandiran, Zagro Singapore Pte Ltd. Singapore  
Mr. Milinkumar V. Sarode, Jawaharlal Darda Institute of Engineering and Technology, India  
Dr. Shamimul Qamar, KSJ Institute of Engineering & Technology, India  
Dr. C. Arun, Anna University, India  
Assist. Prof. M.N.Birje, Basaveshwar Engineering College, India  
Prof. Hamid Reza Naji, Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran  
Assist. Prof. Debasis Giri, Department of Computer Science and Engineering, Haldia Institute of Technology  
Subhabrata Barman, Haldia Institute of Technology, West Bengal  
Mr. M. I. Lali, COMSATS Institute of Information Technology, Islamabad, Pakistan  
Dr. Feroz Khan, Central Institute of Medicinal and Aromatic Plants, Lucknow, India  
Mr. R. Nagendran, Institute of Technology, Coimbatore, Tamilnadu, India  
Mr. Amnach Khawne, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, Thailand

Dr. P. Chakrabarti, Sir Padampat Singhanian University, Udaipur, India  
Mr. Nafiz Imtiaz Bin Hamid, Islamic University of Technology (IUT), Bangladesh.  
Shahab-A. Shamshirband, Islamic Azad University, Chalous, Iran  
Prof. B. Priestly Shan, Anna Univeristy, Tamilnadu, India  
Venkatramreddy Velma, Dept. of Bioinformatics, University of Mississippi Medical Center, Jackson MS USA  
Akshi Kumar, Dept. of Computer Engineering, Delhi Technological University, India  
Dr. Umesh Kumar Singh, Vikram University, Ujjain, India  
Mr. Serguei A. Mokhov, Concordia University, Canada  
Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia  
Dr. Awadhesh Kumar Sharma, Madan Mohan Malviya Engineering College, India  
Mr. Syed R. Rizvi, Analytical Services & Materials, Inc., USA  
Dr. S. Karthik, SNS College of Technology, India  
Mr. Syed Qasim Bukhari, CIMET (Universidad de Granada), Spain  
Mr. A.D.Potgantwar, Pune University, India  
Dr. Himanshu Aggarwal, Punjabi University, India  
Mr. Rajesh Ramachandran, Naipunya Institute of Management and Information Technology, India  
Dr. K.L. Shunmuganathan, R.M.K Engg College , Kavaraipettai ,Chennai  
Dr. Prasant Kumar Pattnaik, KIST, India.  
Dr. Ch. Aswani Kumar, VIT University, India  
Mr. Ijaz Ali Shoukat, King Saud University, Riyadh KSA  
Mr. Arun Kumar, Sir Padam Pat Singhanian University, Udaipur, Rajasthan  
Mr. Muhammad Imran Khan, Universiti Teknologi PETRONAS, Malaysia  
Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA  
Mr. Mohd Zaki Bin Mas'ud, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia  
Prof. Dr. R. Geetharamani, Dept. of Computer Science and Eng., Rajalakshmi Engineering College, India  
Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India  
Dr. S. Abdul Khader Jilani, University of Tabuk, Tabuk, Saudi Arabia  
Mr. Syed Jamal Haider Zaidi, Bahria University, Pakistan  
Dr. N. Devarajan, Government College of Technology, Coimbatore, Tamilnadu, INDIA  
Mr. R. Jagadeesh Kannan, RMK Engineering College, India  
Mr. Deo Prakash, Shri Mata Vaishno Devi University, India  
Mr. Mohammad Abu Naser, Dept. of EEE, IUT, Gazipur, Bangladesh  
Assist. Prof. Prasun Ghosal, Bengal Engineering and Science University, India  
Mr. Md. Golam Kaosar, School of Engineering and Science, Victoria University, Melbourne City, Australia  
Mr. R. Mahammad Shafi, Madanapalle Institute of Technology & Science, India  
Dr. F.Sagayaraj Francis, Pondicherry Engineering College, India  
Dr. Ajay Goel, HIET , Kaithal, India  
Mr. Nayak Sunil Kashibarao, Bahirji Smarak Mahavidyalaya, India  
Mr. Suhas J Manangi, Microsoft India  
Dr. Kalyankar N. V., Yeshwant Mahavidyalaya, Nanded , India  
Dr. K.D. Verma, S.V. College of Post graduate studies & Research, India  
Dr. Amjad Rehman, University Technology Malaysia, Malaysia

Mr. Rachit Garg, L K College, Jalandhar, Punjab  
Mr. J. William, M.A.M college of Engineering, Trichy, Tamilnadu, India  
Prof. Jue-Sam Chou, Nanhua University, College of Science and Technology, Taiwan  
Dr. Thorat S.B., Institute of Technology and Management, India  
Mr. Ajay Prasad, Sir Padampat Singhania University, Udaipur, India  
Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology & Science, India  
Mr. Syed Rafiul Hussain, Ahsanullah University of Science and Technology, Bangladesh  
Mrs Fazeela Tunnisa, Najran University, Kingdom of Saudi Arabia  
Mrs Kavita Taneja, Maharishi Markandeshwar University, Haryana, India  
Mr. Maniyar Shiraz Ahmed, Najran University, Najran, KSA  
Mr. Anand Kumar, AMC Engineering College, Bangalore  
Dr. Rakesh Chandra Gangwar, Beant College of Engg. & Tech., Gurdaspur (Punjab) India  
Dr. V V Rama Prasad, Sree Vidyanikethan Engineering College, India  
Assist. Prof. Neetesh Kumar Gupta, Technocrats Institute of Technology, Bhopal (M.P.), India  
Mr. Ashish Seth, Uttar Pradesh Technical University, Lucknow, UP India  
Dr. V V S S S Balaram, Sreenidhi Institute of Science and Technology, India  
Mr Rahul Bhatia, Lingaya's Institute of Management and Technology, India  
Prof. Niranjana Reddy. P, KITS, Warangal, India  
Prof. Rakesh. Lingappa, Vijetha Institute of Technology, Bangalore, India  
Dr. Mohammed Ali Hussain, Nimra College of Engineering & Technology, Vijayawada, A.P., India  
Dr. A.Srinivasan, MNM Jain Engineering College, Rajiv Gandhi Salai, Thorapakkam, Chennai  
Mr. Rakesh Kumar, M.M. University, Mullana, Ambala, India  
Dr. Lena Khaled, Zarqa Private University, Aman, Jordan  
Ms. Supriya Kapoor, Patni/Lingaya's Institute of Management and Tech., India  
Dr. Tossapon Boongoen, Aberystwyth University, UK  
Dr. Bilal Alatas, Firat University, Turkey  
Assist. Prof. Jyoti Praakash Singh, Academy of Technology, India  
Dr. Ritu Soni, GNG College, India  
Dr. Mahendra Kumar, Sagar Institute of Research & Technology, Bhopal, India.  
Dr. Binod Kumar, Lakshmi Narayan College of Tech.(LNCT) Bhopal India  
Dr. Muzhir Shaban Al-Ani, Amman Arab University Amman – Jordan  
Dr. T.C. Manjunath, ATRIA Institute of Tech, India  
Mr. Muhammad Zakarya, COMSATS Institute of Information Technology (CIIT), Pakistan  
Assist. Prof. Harmunish Taneja, M. M. University, India  
Dr. Chitra Dhawale, SICSR, Model Colony, Pune, India  
Mrs Sankari Muthukaruppan, Nehru Institute of Engineering and Technology, Anna University, India  
Mr. Aaqif Afzaal Abbasi, National University Of Sciences And Technology, Islamabad  
Prof. Ashutosh Kumar Dubey, Trinity Institute of Technology and Research Bhopal, India  
Mr. G. Appasami, Dr. Pauls Engineering College, India  
Mr. M Yasin, National University of Science and Tech, Karachi (NUST), Pakistan  
Mr. Yaser Miaji, University Utara Malaysia, Malaysia  
Mr. Shah Ahsanul Haque, International Islamic University Chittagong (IIUC), Bangladesh

Prof. (Dr) Syed Abdul Sattar, Royal Institute of Technology & Science, India  
Dr. S. Sasikumar, Roever Engineering College  
Assist. Prof. Monit Kapoor, Maharishi Markandeshwar University, India  
Mr. Nwaocha Vivian O, National Open University of Nigeria  
Dr. M. S. Vijaya, GR Govindarajulu School of Applied Computer Technology, India  
Assist. Prof. Chakresh Kumar, Manav Rachna International University, India  
Mr. Kunal Chadha , R&D Software Engineer, Gemalto, Singapore  
Mr. Mueen Uddin, Universiti Teknologi Malaysia, UTM , Malaysia  
Dr. Dhuha Basheer abdullah, Mosul university, Iraq  
Mr. S. Audithan, Annamalai University, India  
Prof. Vijay K Chaudhari, Technocrats Institute of Technology , India  
Associate Prof. Mohd Ilyas Khan, Technocrats Institute of Technology , India  
Dr. Vu Thanh Nguyen, University of Information Technology, HoChiMinh City, VietNam  
Assist. Prof. Anand Sharma, MITS, Lakshmandarh, Sikar, Rajasthan, India  
Prof. T V Narayana Rao, HITAM Engineering college, Hyderabad  
Mr. Deepak Gour, Sir Padampat Singhania University, India  
Assist. Prof. Amutharaj Joyson, Kalasalingam University, India  
Mr. Ali Balador, Islamic Azad University, Iran  
Mr. Mohit Jain, Maharaja Surajmal Institute of Technology, India  
Mr. Dilip Kumar Sharma, GLA Institute of Technology & Management, India  
Dr. Debojyoti Mitra, Sir padampat Singhania University, India  
Dr. Ali Dehghantanha, Asia-Pacific University College of Technology and Innovation, Malaysia  
Mr. Zhao Zhang, City University of Hong Kong, China  
Prof. S.P. Setty, A.U. College of Engineering, India  
Prof. Patel Rakeshkumar Kantilal, Sankalchand Patel College of Engineering, India  
Mr. Biswajit Bhowmik, Bengal College of Engineering & Technology, India  
Mr. Manoj Gupta, Apex Institute of Engineering & Technology, India  
Assist. Prof. Ajay Sharma, Raj Kumar Goel Institute Of Technology, India  
Assist. Prof. Ramveer Singh, Raj Kumar Goel Institute of Technology, India  
Dr. Hanan Elazhary, Electronics Research Institute, Egypt  
Dr. Hosam I. Faiq, USM, Malaysia  
Prof. Dipti D. Patil, MAEER's MIT College of Engg. & Tech, Pune, India  
Assist. Prof. Devendra Chack, BCT Kumaon engineering College Dwarahat Almora, India  
Prof. Manpreet Singh, M. M. Engg. College, M. M. University, India  
Assist. Prof. M. Sadiq ali Khan, University of Karachi, Pakistan  
Mr. Prasad S. Halgaonkar, MIT - College of Engineering, Pune, India  
Dr. Imran Ghani, Universiti Teknologi Malaysia, Malaysia  
Prof. Varun Kumar Kakar, Kumaon Engineering College, Dwarahat, India  
Assist. Prof. Nisheeth Joshi, Apaji Institute, Banasthali University, Rajasthan, India  
Associate Prof. Kunwar S. Vaisla, VCT Kumaon Engineering College, India  
Prof Anupam Choudhary, Bhilai School Of Engg.,Bhilai (C.G.),India  
Mr. Divya Prakash Shrivastava, Al Jabal Al garbi University, Zawya, Libya

Associate Prof. Dr. V. Radha, Avinashilingam Deemed university for women, Coimbatore.  
Dr. Kasarapu Ramani, JNT University, Anantapur, India  
Dr. Anuraag Awasthi, Jayoti Vidyapeeth Womens University, India  
Dr. C G Ravichandran, R V S College of Engineering and Technology, India  
Dr. Mohamed A. Deriche, King Fahd University of Petroleum and Minerals, Saudi Arabia  
Mr. Abbas Karimi, Universiti Putra Malaysia, Malaysia  
Mr. Amit Kumar, Jaypee University of Engg. and Tech., India  
Dr. Nikolai Stoianov, Defense Institute, Bulgaria  
Assist. Prof. S. Ranichandra, KSR College of Arts and Science, Tiruchencode  
Mr. T.K.P. Rajagopal, Diamond Horse International Pvt Ltd, India  
Dr. Md. Ekramul Hamid, Rajshahi University, Bangladesh  
Mr. Hemanta Kumar Kalita , TATA Consultancy Services (TCS), India  
Dr. Messaouda Azzouzi, Ziane Achour University of Djelfa, Algeria  
Prof. (Dr.) Juan Jose Martinez Castillo, "Gran Mariscal de Ayacucho" University and Acantelys research Group, Venezuela  
Dr. Jatinderkumar R. Saini, Narmada College of Computer Application, India  
Dr. Babak Bashari Rad, University Technology of Malaysia, Malaysia  
Dr. Nighat Mir, Effat University, Saudi Arabia  
Prof. (Dr.) G.M.Nasira, Sasurie College of Engineering, India  
Mr. Varun Mittal, Gemalto Pte Ltd, Singapore  
Assist. Prof. Mrs P. Banumathi, Kathir College Of Engineering, Coimbatore  
Assist. Prof. Quan Yuan, University of Wisconsin-Stevens Point, US  
Dr. Pranam Paul, Narula Institute of Technology, Agarpara, West Bengal, India  
Assist. Prof. J. Ramkumar, V.L.B Janakiammal college of Arts & Science, India  
Mr. P. Sivakumar, Anna university, Chennai, India  
Mr. Md. Humayun Kabir Biswas, King Khalid University, Kingdom of Saudi Arabia  
Mr. Mayank Singh, J.P. Institute of Engg & Technology, Meerut, India  
HJ. Kamaruzaman Jusoff, Universiti Putra Malaysia  
Mr. Nikhil Patrick Lobo, CADES, India  
Dr. Amit Wason, Rayat-Bahra Institute of Engineering & Boi-Technology, India  
Dr. Rajesh Shrivastava, Govt. Benazir Science & Commerce College, Bhopal, India  
Assist. Prof. Vishal Bharti, DCE, Gurgaon  
Mrs. Sunita Bansal, Birla Institute of Technology & Science, India  
Dr. R. Sudhakar, Dr.Mahalingam college of Engineering and Technology, India  
Dr. Amit Kumar Garg, Shri Mata Vaishno Devi University, Katra(J&K), India  
Assist. Prof. Raj Gaurang Tiwari, AZAD Institute of Engineering and Technology, India  
Mr. Hamed Taherdoost, Tehran, Iran  
Mr. Amin Daneshmand Malayeri, YRC, IAU, Malayer Branch, Iran  
Mr. Shantanu Pal, University of Calcutta, India  
Dr. Terry H. Walcott, E-Promag Consultancy Group, United Kingdom  
Dr. Ezekiel U OKIKE, University of Ibadan, Nigeria  
Mr. P. Mahalingam, Caledonian College of Engineering, Oman

Dr. Mahmoud M. A. Abd Ellatif, Mansoura University, Egypt  
Prof. Kunwar S. Vaisla, BCT Kumaon Engineering College, India  
Prof. Mahesh H. Panchal, Kalol Institute of Technology & Research Centre, India  
Mr. Muhammad Asad, Technical University of Munich, Germany  
Mr. AliReza Shams Shafigh, Azad Islamic university, Iran  
Prof. S. V. Nagaraj, RMK Engineering College, India  
Mr. Ashikali M Hasan, Senior Researcher, CelNet security, India  
Dr. Adnan Shahid Khan, University Technology Malaysia, Malaysia  
Mr. Prakash Gajanan Burade, Nagpur University/ITM college of engg, Nagpur, India  
Dr. Jagdish B.Helonde, Nagpur University/ITM college of engg, Nagpur, India  
Professor, Doctor BOUHORMA Mohammed, Univertsity Abdelmalek Essaadi, Morocco  
Mr. K. Thirumalaivasan, Pondicherry Engg. College, India  
Mr. Umbarkar Anantkumar Janardan, Walchand College of Engineering, India  
Mr. Ashish Chaurasia, Gyan Ganga Institute of Technology & Sciences, India  
Mr. Sunil Taneja, Kurukshetra University, India  
Mr. Fauzi Adi Rafrastara, Dian Nuswantoro University, Indonesia  
Dr. Yaduvir Singh, Thapar University, India  
Dr. Ioannis V. Koskosas, University of Western Macedonia, Greece  
Dr. Vasantha Kalyani David, Avinashilingam University for women, Coimbatore  
Dr. Ahmed Mansour Manasrah, Universiti Sains Malaysia, Malaysia  
Miss. Nazanin Sadat Kazazi, University Technology Malaysia, Malaysia  
Mr. Saeed Rasouli Heikalabad, Islamic Azad University - Tabriz Branch, Iran  
Assoc. Prof. Dharendra Mishra, SVKM's NMIMS University, India  
Prof. Shapoor Zarei, UAE Inventors Association, UAE  
Prof. B.Raja Sarath Kumar, Lenora College of Engineering, India  
Dr. Bashir Alam, Jamia millia Islamia, Delhi, India  
Prof. Anant J Umbarkar, Walchand College of Engg., India  
Assist. Prof. B. Bharathi, Sathyabama University, India  
Dr. Fokrul Alom Mazarbhuiya, King Khalid University, Saudi Arabia  
Prof. T.S.Jeyali Laseeth, Anna University of Technology, Tirunelveli, India  
Dr. M. Balraju, Jawahar Lal Nehru Technological University Hyderabad, India  
Dr. Vijayalakshmi M. N., R.V.College of Engineering, Bangalore  
Prof. Walid Moudani, Lebanese University, Lebanon  
Dr. Saurabh Pal, VBS Purvanchal University, Jaunpur, India  
Associate Prof. Suneet Chaudhary, Dehradun Institute of Technology, India  
Associate Prof. Dr. Manuj Darbari, BBD University, India  
Ms. Prema Selvaraj, K.S.R College of Arts and Science, India  
Assist. Prof. Ms.S.Sasikala, KSR College of Arts & Science, India  
Mr. Sukhvinder Singh Deora, NC Institute of Computer Sciences, India  
Dr. Abhay Bansal, Amity School of Engineering & Technology, India  
Ms. Sumita Mishra, Amity School of Engineering and Technology, India  
Professor S. Viswanadha Raju, JNT University Hyderabad, India

Mr. Asghar Shahrzad Khashandarag, Islamic Azad University Tabriz Branch, India  
Mr. Manoj Sharma, Panipat Institute of Engg. & Technology, India  
Mr. Shakeel Ahmed, King Faisal University, Saudi Arabia  
Dr. Mohamed Ali Mahjoub, Institute of Engineer of Monastir, Tunisia  
Mr. Adri Jovin J.J., SriGuru Institute of Technology, India  
Dr. Sukumar Senthilkumar, Universiti Sains Malaysia, Malaysia  
Mr. Rakesh Bharati, Dehradun Institute of Technology Dehradun, India  
Mr. Shervan Fekri Ershad, Shiraz International University, Iran  
Mr. Md. Safiqul Islam, Daffodil International University, Bangladesh  
Mr. Mahmudul Hasan, Daffodil International University, Bangladesh  
Prof. Mandakini Tayade, UIT, RGTU, Bhopal, India  
Ms. Sarla More, UIT, RGTU, Bhopal, India  
Mr. Tushar Hrishikesh Jaware, R.C. Patel Institute of Technology, Shirpur, India  
Ms. C. Divya, Dr G R Damodaran College of Science, Coimbatore, India  
Mr. Fahimuddin Shaik, Annamacharya Institute of Technology & Sciences, India  
Dr. M. N. Giri Prasad, JNTUCE,Pulivendula, A.P., India  
Assist. Prof. Chintan M Bhatt, Charotar University of Science And Technology, India  
Prof. Sahista Machchhar, Marwadi Education Foundation's Group of institutions, India  
Assist. Prof. Navnish Goel, S. D. College Of Enginnering & Technology, India  
Mr. Khaja Kamaluddin, Sirt University, Sirt, Libya  
Mr. Mohammad Zaidul Karim, Daffodil International, Bangladesh  
Mr. M. Vijayakumar, KSR College of Engineering, Tiruchengode, India  
Mr. S. A. Ahsan Rajon, Khulna University, Bangladesh  
Dr. Muhammad Mohsin Nazir, LCW University Lahore, Pakistan  
Mr. Mohammad Asadul Hoque, University of Alabama, USA  
Mr. P.V.Sarathchand, Indur Institute of Engineering and Technology, India  
Mr. Durgesh Samadhiya, Chung Hua University, Taiwan  
Dr Venu Kuthadi, University of Johannesburg, Johannesburg, RSA  
Dr. (Er) Jasvir Singh, Guru Nanak Dev University, Amritsar, Punjab, India  
Mr. Jasmin Cosic, Min. of the Interior of Una-sana canton, B&H, Bosnia and Herzegovina  
Dr S. Rajalakshmi, Botho College, South Africa  
Dr. Mohamed Sarrab, De Montfort University, UK  
Mr. Basappa B. Kodada, Canara Engineering College, India  
Assist. Prof. K. Ramana, Annamacharya Institute of Technology and Sciences, India  
Dr. Ashu Gupta, Apeejay Institute of Management, Jalandhar, India  
Assist. Prof. Shaik Rasool, Shadan College of Engineering & Technology, India  
Assist. Prof. K. Suresh, Annamacharya Institute of Tech & Sci. Rajampet, AP, India  
Dr . G. Singaravel, K.S.R. College of Engineering, India  
Dr B. G. Geetha, K.S.R. College of Engineering, India  
Assist. Prof. Kavita Choudhary, ITM University, Gurgaon  
Dr. Mehrdad Jalali, Azad University, Mashhad, Iran  
Megha Goel, Shamli Institute of Engineering and Technology, Shamli, India

Mr. Chi-Hua Chen, Institute of Information Management, National Chiao-Tung University, Taiwan (R.O.C.)  
Assoc. Prof. A. Rajendran, RVS College of Engineering and Technology, India  
Assist. Prof. S. Jaganathan, RVS College of Engineering and Technology, India  
Assoc. Prof. (Dr.) A S N Chakravarthy, JNTUK University College of Engineering Vizianagaram (State University)  
Assist. Prof. Deepshikha Patel, Technocrat Institute of Technology, India  
Assist. Prof. Maram Balajee, GMRIT, India  
Assist. Prof. Monika Bhatnagar, TIT, India  
Prof. Gaurang Panchal, Charotar University of Science & Technology, India  
Prof. Anand K. Tripathi, Computer Society of India  
Prof. Jyoti Chaudhary, High Performance Computing Research Lab, India  
Assist. Prof. Supriya Raheja, ITM University, India  
Dr. Pankaj Gupta, Microsoft Corporation, U.S.A.  
Assist. Prof. Panchamukesh Chandaka, Hyderabad Institute of Tech. & Management, India  
Prof. Mohan H.S, SJB Institute Of Technology, India  
Mr. Hossein Malekinezhad, Islamic Azad University, Iran  
Mr. Zatin Gupta, Universti Malaysia, Malaysia  
Assist. Prof. Amit Chauhan, Phonics Group of Institutions, India  
Assist. Prof. Ajal A. J., METS School Of Engineering, India  
Mrs. Omowunmi Omobola Adeyemo, University of Ibadan, Nigeria  
Dr. Bharat Bhushan Agarwal, I.F.T.M. University, India  
Md. Nazrul Islam, University of Western Ontario, Canada  
Tushar Kanti, L.N.C.T, Bhopal, India  
Er. Aumreesh Kumar Saxena, SIRTs College Bhopal, India  
Mr. Mohammad Monirul Islam, Daffodil International University, Bangladesh  
Dr. Kashif Nisar, University Utara Malaysia, Malaysia  
Dr. Wei Zheng, Rutgers Univ/ A10 Networks, USA  
Associate Prof. Rituraj Jain, Vyas Institute of Engg & Tech, Jodhpur – Rajasthan  
Assist. Prof. Apoorvi Sood, I.T.M. University, India  
Dr. Kayhan Zrar Ghafoor, University Technology Malaysia, Malaysia  
Mr. Swapnil Soner, Truba Institute College of Engineering & Technology, Indore, India  
Ms. Yogita Gigras, I.T.M. University, India  
Associate Prof. Neelima Sadineni, Pydha Engineering College, India Pydha Engineering College  
Assist. Prof. K. Deepika Rani, HITAM, Hyderabad  
Ms. Shikha Maheshwari, Jaipur Engineering College & Research Centre, India  
Prof. Dr V S Giridhar Akula, Avanthi's Scientific Tech. & Research Academy, Hyderabad  
Prof. Dr.S.Saravanan, Muthayammal Engineering College, India  
Mr. Mehdi Golsorkhatabar Amiri, Islamic Azad University, Iran  
Prof. Amit Sadanand Savyanavar, MITCOE, Pune, India  
Assist. Prof. P.Oliver Jayaprakash, Anna University, Chennai  
Assist. Prof. Ms. Sujata, ITM University, Gurgaon, India  
Dr. Asoke Nath, St. Xavier's College, India

Mr. Masoud Rafighi, Islamic Azad University, Iran  
Assist. Prof. RamBabu Pemula, NIMRA College of Engineering & Technology, India  
Assist. Prof. Ms Rita Chhikara, ITM University, Gurgaon, India  
Mr. Sandeep Maan, Government Post Graduate College, India  
Prof. Dr. S. Muralidharan, Mepco Schlenk Engineering College, India  
Associate Prof. T.V.Sai Krishna, QIS College of Engineering and Technology, India  
Mr. R. Balu, Bharathiar University, Coimbatore, India  
Assist. Prof. Shekhar. R, Dr.SM College of Engineering, India  
Prof. P. Senthilkumar, Vivekanandha Institute of Engineering and Technology for Woman, India  
Mr. M. Kamarajan, PSNA College of Engineering & Technology, India  
Dr. Angajala Srinivasa Rao, Jawaharlal Nehru Technical University, India  
Assist. Prof. C. Venkatesh, A.I.T.S, Rajampet, India  
Mr. Afshin Rezakhani Roozbahani, Ayatollah Boroujerdi University, Iran  
Mr. Laxmi chand, SCTL, Noida, India  
Dr. Dr. Abdul Hannan, Vivekanand College, Aurangabad  
Prof. Mahesh Panchal, KITRC, Gujarat  
Dr. A. Subramani, K.S.R. College of Engineering, Tiruchengode  
Assist. Prof. Prakash M, Rajalakshmi Engineering College, Chennai, India  
Assist. Prof. Akhilesh K Sharma, Sir Padampat Singhania University, India  
Ms. Varsha Sahni, Guru Nanak Dev Engineering College, Ludhiana, India  
Associate Prof. Trilochan Rout, NM Institute of Engineering and Technology, India  
Mr. Srikanta Kumar Mohapatra, NMIET, Orissa, India  
Mr. Waqas Haider Bangyal, Iqra University Islamabad, Pakistan  
Dr. S. Vijayaragavan, Christ College of Engineering and Technology, Pondicherry, India  
Prof. Elboukhari Mohamed, University Mohammed First, Oujda, Morocco  
Dr. Muhammad Asif Khan, King Faisal University, Saudi Arabia  
Dr. Nagy Ramadan Darwish Omran, Cairo University, Egypt.  
Assistant Prof. Anand Nayyar, KCL Institute of Management and Technology, India  
Mr. G. Premsankar, Ericsson, India  
Assist. Prof. T. Hemalatha, VELS University, India  
Prof. Tejaswini Apte, University of Pune, India  
Dr. Edmund Ng Giap Weng, Universiti Malaysia Sarawak, Malaysia  
Mr. Mahdi Nouri, Iran University of Science and Technology, Iran  
Associate Prof. S. Asif Hussain, Annamacharya Institute of technology & Sciences, India  
Mrs. Kavita Pabreja, Maharaja Surajmal Institute (an affiliate of GGSIP University), India  
Mr. Vorugunti Chandra Sekhar, DA-IICT, India  
Mr. Muhammad Najmi Ahmad Zabidi, Universiti Teknologi Malaysia, Malaysia  
Dr. Aderemi A. Atayero, Covenant University, Nigeria  
Assist. Prof. Osama Sohaib, Balochistan University of Information Technology, Pakistan  
Assist. Prof. K. Suresh, Annamacharya Institute of Technology and Sciences, India  
Mr. Hassen Mohammed Abdulllah Alsafi, International Islamic University Malaysia (IIUM) Malaysia  
Mr. Robail Yasrab, Virtual University of Pakistan, Pakistan

Mr. R. Balu, Bharathiar University, Coimbatore, India  
Prof. Anand Nayyar, KCL Institute of Management and Technology, Jalandhar  
Assoc. Prof. Vivek S Deshpande, MIT College of Engineering, India  
Prof. K. Saravanan, Anna university Coimbatore, India  
Dr. Ravendra Singh, MJP Rohilkhand University, Bareilly, India  
Mr. V. Mathivanan, IBRA College of Technology, Sultanate of OMAN  
Assoc. Prof. S. Asif Hussain, AITS, India  
Assist. Prof. C. Venkatesh, AITS, India  
Mr. Sami Ulhaq, SZABIST Islamabad, Pakistan  
Dr. B. Justus Rabi, Institute of Science & Technology, India  
Mr. Anuj Kumar Yadav, Dehradun Institute of technology, India  
Mr. Alejandro Mosquera, University of Alicante, Spain  
Assist. Prof. Arjun Singh, Sir Padampat Singhania University (SPSU), Udaipur, India  
Dr. Smriti Agrawal, JB Institute of Engineering and Technology, Hyderabad  
Assist. Prof. Swathi Sambangi, Visakha Institute of Engineering and Technology, India  
Ms. Prabhjot Kaur, Guru Gobind Singh Indraprastha University, India  
Mrs. Samaher AL-Hothali, Yanbu University College, Saudi Arabia  
Prof. Rajneeshkaur Bedi, MIT College of Engineering, Pune, India  
Mr. Hassen Mohammed Abdullallah Alsafi, International Islamic University Malaysia (IIUM)  
Dr. Wei Zhang, Amazon.com, Seattle, WA, USA  
Mr. B. Santhosh Kumar, C S I College of Engineering, Tamil Nadu  
Dr. K. Reji Kumar, , N S S College, Pandalam, India  
Assoc. Prof. K. Seshadri Sastry, EIILM University, India  
Mr. Kai Pan, UNC Charlotte, USA  
Mr. Ruikar Sachin, SGGSIET, India  
Prof. (Dr.) Vinodani Katiyar, Sri Ramswaroop Memorial University, India  
Assoc. Prof., M. Giri, Sreenivasa Institute of Technology and Management Studies, India  
Assoc. Prof. Labib Francis Gergis, Misr Academy for Engineering and Technology (MET), Egypt  
Assist. Prof. Amanpreet Kaur, ITM University, India  
Assist. Prof. Anand Singh Rajawat, Shri Vaishnav Institute of Technology & Science, Indore  
Mrs. Hadeel Saleh Haj Aliwi, Universiti Sains Malaysia (USM), Malaysia  
Dr. Abhay Bansal, Amity University, India  
Dr. Mohammad A. Mezher, Fahad Bin Sultan University, KSA  
Assist. Prof. Nidhi Arora, M.C.A. Institute, India  
Prof. Dr. P. Suresh, Karpagam College of Engineering, Coimbatore, India  
Dr. Kannan Balasubramanian, Mepco Schlenk Engineering College, India  
Dr. S. Sankara Gomathi, Panimalar Engineering college, India  
Prof. Anil kumar Suthar, Gujarat Technological University, L.C. Institute of Technology, India  
Assist. Prof. R. Hubert Rajan, NOORUL ISLAM UNIVERSITY, India  
Assist. Prof. Dr. Jyoti Mahajan, College of Engineering & Technology  
Assist. Prof. Homam Reda El-Taj, College of Network Engineering, Saudi Arabia & Malaysia  
Mr. Bijan Paul, Shahjalal University of Science & Technology, Bangladesh

Assoc. Prof. Dr. Ch V Phani Krishna, KL University, India  
Dr. Vishal Bhatnagar, Ambedkar Institute of Advanced Communication Technologies & Research, India  
Dr. Lamri LAOUAMER, Al Qassim University, Dept. Info. Systems & European University of Brittany, Dept.  
Computer Science, UBO, Brest, France  
Prof. Ashish Babanrao Sasankar, G.H.Raisoni Institute Of Information Technology, India  
Prof. Pawan Kumar Goel, Shamli Institute of Engineering and Technology, India  
Mr. Ram Kumar Singh, S.V Subharti University, India  
Assistant Prof. Sunish Kumar O S, Amaljiyothi College of Engineering, India  
Dr Sanjay Bhargava, Banasthali University, India  
Mr. Pankaj S. Kulkarni, AVEW's Shatabdi Institute of Technology, India  
Mr. Roohollah Etemadi, Islamic Azad University, Iran  
Mr. Oloruntoyin Sefiu Taiwo, Emmanuel Alayande College Of Education, Nigeria  
Mr. Sumit Goyal, National Dairy Research Institute, India  
Mr Jaswinder Singh Dilawari, Geeta Engineering College, India  
Prof. Raghuraj Singh, Harcourt Butler Technological Institute, Kanpur  
Dr. S.K. Mahendran, Anna University, Chennai, India  
Dr. Amit Wason, Hindustan Institute of Technology & Management, Punjab  
Dr. Ashu Gupta, Apeejay Institute of Management, India  
Assist. Prof. D. Asir Antony Gnana Singh, M.I.E.T Engineering College, India  
Mrs Mina Farmanbar, Eastern Mediterranean University, Famagusta, North Cyprus  
Mr. Maram Balajee, GMR Institute of Technology, India  
Mr. Moiz S. Ansari, Isra University, Hyderabad, Pakistan  
Mr. Adebayo, Olawale Surajudeen, Federal University of Technology Minna, Nigeria  
Mr. Jasvir Singh, University College Of Engg., India  
Mr. Vivek Tiwari, MANIT, Bhopal, India  
Assoc. Prof. R. Navaneethakrishnan, Bharathiyar College of Engineering and Technology, India  
Mr. Somdip Dey, St. Xavier's College, Kolkata, India  
Mr. Souleymane Balla-Arabé, Xi'an University of Electronic Science and Technology, China  
Mr. Mahabub Alam, Rajshahi University of Engineering and Technology, Bangladesh  
Mr. Sathyapraksh P., S.K.P Engineering College, India  
Dr. N. Karthikeyan, SNS College of Engineering, Anna University, India  
Dr. Binod Kumar, JSPM's, Jayawant Technical Campus, Pune, India  
Assoc. Prof. Dinesh Goyal, Suresh Gyan Vihar University, India  
Mr. Md. Abdul Ahad, K L University, India  
Mr. Vikas Bajpai, The LNM IIT, India  
Dr. Manish Kumar Anand, Salesforce (R & D Analytics), San Francisco, USA  
Assist. Prof. Dheeraj Murari, Kumaon Engineering College, India  
Assoc. Prof. Dr. A. Muthukumaravel, VELS University, Chennai  
Mr. A. Siles Balasingh, St. Joseph University in Tanzania, Tanzania  
Mr. Ravindra Daga Badgujar, R C Patel Institute of Technology, India  
Dr. Preeti Khanna, SVKM's NMIMS, School of Business Management, India  
Mr. Kumar Dayanand, Cambridge Institute of Technology, India

Dr. Syed Asif Ali, SMI University Karachi, Pakistan  
Prof. Pallvi Pandit, Himachal Pradesh University, India  
Mr. Ricardo Verschueren, University of Gloucestershire, UK  
Assist. Prof. Mamta Juneja, University Institute of Engineering and Technology, Panjab University, India  
Assoc. Prof. P. Surendra Varma, NRI Institute of Technology, JNTU Kakinada, India  
Assist. Prof. Gaurav Shrivastava, RGPV / SVITS Indore, India  
Dr. S. Sumathi, Anna University, India  
Assist. Prof. Ankita M. Kapadia, Charotar University of Science and Technology, India  
Mr. Deepak Kumar, Indian Institute of Technology (BHU), India  
Dr. Dr. Rajan Gupta, GGSIP University, New Delhi, India  
Assist. Prof M. Anand Kumar, Karpagam University, Coimbatore, India  
Mr. Mr Arshad Mansoor, Pakistan Aeronautical Complex  
Mr. Kapil Kumar Gupta, Ansal Institute of Technology and Management, India  
Dr. Neeraj Tomer, SINE International Institute of Technology, Jaipur, India  
Assist. Prof. Trunal J. Patel, C.G.Patel Institute of Technology, Uka Tarsadia University, Bardoli, Surat  
Mr. Sivakumar, Codework solutions, India  
Mr. Mohammad Sadegh Mirzaei, PGNR Company, Iran  
Dr. Gerard G. Dumancas, Oklahoma Medical Research Foundation, USA  
Mr. Varadala Sridhar, Varadhman College Engineering College, Affiliated To JNTU, Hyderabad  
Assist. Prof. Manoj Dhawan, SVITS, Indore  
Assoc. Prof. Chitreshh Banerjee, Suresh Gyan Vihar University, Jaipur, India  
Dr. S. Santhi, SCSVMV University, India  
Mr. Davood Mohammadi Souran, Ministry of Energy of Iran, Iran  
Mr. Shamim Ahmed, Bangladesh University of Business and Technology, Bangladesh  
Mr. Sandeep Reddivari, Mississippi State University, USA  
Assoc. Prof. Ousmane Thiare, Gaston Berger University, Senegal  
Dr. Hazra Imran, Athabasca University, Canada  
Dr. Setu Kumar Chaturvedi, Technocrats Institute of Technology, Bhopal, India  
Mr. Mohd Dilshad Ansari, Jaypee University of Information Technology, India  
Ms. Jaspreet Kaur, Distance Education LPU, India  
Dr. D. Nagarajan, Salalah College of Technology, Sultanate of Oman  
Dr. K.V.N.R.Sai Krishna, S.V.R.M. College, India  
Mr. Himanshu Pareek, Center for Development of Advanced Computing (CDAC), India  
Mr. Khaldi Amine, Badji Mokhtar University, Algeria  
Mr. Mohammad Sadegh Mirzaei, Scientific Applied University, Iran  
Assist. Prof. Khyati Chaudhary, Ram-eesh Institute of Engg. & Technology, India  
Mr. Sanjay Agal, Pacific College of Engineering Udaipur, India  
Mr. Abdul Mateen Ansari, King Khalid University, Saudi Arabia  
Dr. H.S. Behera, Veer Surendra Sai University of Technology (VSSUT), India  
Dr. Shrikant Tiwari, Shri Shankaracharya Group of Institutions (SSGI), India  
Prof. Ganesh B. Regulwar, Shri Shankarprasad Agnihotri College of Engg, India  
Prof. Pinnamaneni Bhanu Prasad, Matrix vision GmbH, Germany

Dr. Shrikant Tiwari, Shri Shankaracharya Technical Campus (SSTC), India  
Dr. Siddesh G.K., : Dayananada Sagar College of Engineering, Bangalore, India  
Dr. Nadir Bouchama, CERIST Research Center, Algeria  
Dr. R. Sathishkumar, Sri Venkateswara College of Engineering, India  
Assistant Prof (Dr.) Mohamed Moussaoui, Abdelmalek Essaadi University, Morocco  
Dr. S. Malathi, Panimalar Engineering College, Chennai, India  
Dr. V. Subedha, Panimalar Institute of Technology, Chennai, India  
Dr. Prashant Panse, Swami Vivekanand College of Engineering, Indore, India  
Dr. Hamza Aldabbas, Al-Balqa'a Applied University, Jordan  
Dr. G. Rasitha Banu, Vel's University, Chennai  
Dr. V. D. Ambeth Kumar, Panimalar Engineering College, Chennai  
Prof. Anuranjan Misra, Bhagwant Institute of Technology, Ghaziabad, India  
Ms. U. Sinthuja, PSG college of arts & science, India  
Dr. Ehsan Saradar Torshizi, Urmia University, Iran  
Dr. Shamneesh Sharma, APG Shimla University, Shimla (H.P.), India  
Assistant Prof. A. S. Syed Navaz, Muthayammal College of Arts & Science, India  
Assistant Prof. Ranjit Panigrahi, Sikkim Manipal Institute of Technology, Majitar, Sikkim  
Dr. Khaled Eskaf, Arab Academy for Science ,Technology & Maritime Transportation, Egypt  
Dr. Nishant Gupta, University of Jammu, India  
Assistant Prof. Nagarajan Sankaran, Annamalai University, Chidambaram, Tamilnadu, India  
Assistant Prof. Tribikram Pradhan, Manipal Institute of Technology, India  
Dr. Nasser Lotfi, Eastern Mediterranean University, Northern Cyprus  
Dr. R. Manavalan, K S Rangasamy college of Arts and Science, Tamilnadu, India  
Assistant Prof. P. Krishna Sankar, K S Rangasamy college of Arts and Science, Tamilnadu, India  
Dr. Rahul Malik, Cisco Systems, USA  
Dr. S. C. Lingareddy, ALPHA College of Engineering, India  
Assistant Prof. Mohammed Shuaib, Interl University, Lucknow, India  
Dr. Sachin Yele, Sanghvi Institute of Management & Science, India  
Dr. T. Thambidurai, Sun Univercell, Singapore  
Prof. Anandkumar Telang, BKIT, India  
Assistant Prof. R. Poorvadevi, SCSVMV University, India  
Dr Uttam Mande, Gitam University, India  
Dr. Poornima Girish Naik, Shahu Institute of Business Education and Research (SIBER), India  
Prof. Md. Abu Kausar, Jaipur National University, Jaipur, India  
Dr. Mohammed Zuber, AISECT University, India  
Prof. Kalum Priyanath Udagepola, King Abdulaziz University, Saudi Arabia  
Dr. K. R. Ananth, Velalar College of Engineering and Technology, India  
Assistant Prof. Sanjay Sharma, Roorkee Engineering & Management Institute Shamli (U.P), India  
Assistant Prof. Panem Charan Arur, Priyadarshini Institute of Technology, India  
Dr. Ashwak Mahmood muhsen alabaichi, Karbala University / College of Science, Iraq  
Dr. Urmila Shrawankar, G H Raison College of Engineering, Nagpur (MS), India  
Dr. Krishan Kumar Paliwal, Panipat Institute of Engineering & Technology, India

Dr. Mukesh Negi, Tech Mahindra, India  
Dr. Anuj Kumar Singh, Amity University Gurgaon, India  
Dr. Babar Shah, Gyeongsang National University, South Korea  
Assistant Prof. Jayprakash Upadhyay, SRI-TECH Jabalpur, India  
Assistant Prof. Varadala Sridhar, Vidya Jyothi Institute of Technology, India  
Assistant Prof. Parameshachari B D, KSIT, Bangalore, India  
Assistant Prof. Ankit Garg, Amity University, Haryana, India  
Assistant Prof. Rajashe Karappa, SDMCET, Karnataka, India  
Assistant Prof. Varun Jasuja, GNIT, India  
Assistant Prof. Sonal Honale, Abha Gaikwad Patil College of Engineering Nagpur, India  
Dr. Pooja Choudhary, CT Group of Institutions, NIT Jalandhar, India  
Dr. Faouzi Hidoussi, UHL Batna, Algeria  
Dr. Naseer Ali Husieen, Wasit University, Iraq  
Assistant Prof. Vinod Kumar Shukla, Amity University, Dubai  
Dr. Ahmed Farouk Metwaly, K L University  
Mr. Mohammed Noaman Murad, Cihan University, Iraq  
Dr. Suxing Liu, Arkansas State University, USA  
Dr. M. Gomathi, Velalar College of Engineering and Technology, India  
Assistant Prof. Sumardiono, College PGRI Blitar, Indonesia  
Dr. Latika Kharb, Jagan Institute of Management Studies (JIMS), Delhi, India  
Associate Prof. S. Raja, Pauls College of Engineering and Technology, Tamilnadu, India  
Assistant Prof. Seyed Reza Pakize, Shahid Sani High School, Iran  
Dr. Thiyagu Nagaraj, University-INOUE, India  
Assistant Prof. Noreen Sarai, Harare Institute of Technology, Zimbabwe  
Assistant Prof. Gajanand Sharma, Suresh Gyan Vihar University Jaipur, Rajasthan, India  
Assistant Prof. Mapari Vikas Prakash, Siddhant COE, Sudumbare, Pune, India  
Dr. Devesh Katiyar, Shri Ramswaroop Memorial University, India  
Dr. Shenshen Liang, University of California, Santa Cruz, US  
Assistant Prof. Mohammad Abu Omar, Limkokwing University of Creative Technology- Malaysia  
Mr. Snehasis Banerjee, Tata Consultancy Services, India  
Assistant Prof. Kibona Lusekelo, Ruaha Catholic University (RUCU), Tanzania  
Assistant Prof. Adib Kabir Chowdhury, University College Technology Sarawak, Malaysia

# CALL FOR PAPERS

## International Journal of Computer Science and Information Security

**IJCSIS 2015**

**ISSN: 1947-5500**

<http://sites.google.com/site/ijcsis/>

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

### ***Track A: Security***

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity  
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc.), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

### ***Track B: Computer Science***

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embedded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail [ijcsiseditor@gmail.com](mailto:ijcsiseditor@gmail.com). Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://sites.google.com/site/ijcsis/authors-notes> .



**© IJCSIS PUBLICATION 2015**

**ISSN 1947 5500**

**<http://sites.google.com/site/ijcsis/>**