In the Name of God

Journal of
Information Systems & Telecommunication
Vol. 2, No. 1, January-March 2014, Serial Number 5

Indexed in:
- Research Institute for Information and Communication Technology   www.ictrc.ir
- Islamic World Science Citation Center (ISC)   www.isc.gov.ir
- Scientific Information Database (SID)   www.sid.ir
- Regional Information Center for Science and Technology (RICeST)   www.srlst.com
- Magiran   www.magiran.com
- CIVILICA   www.civilica.com

Acknowledgement

Table of Contents

A Learning Automata Approach to Cooperative Particle Swarm Optimizer

Mohammad Hasanzadeh*
Computer Engineering and Information Technology Department, Amirkabir University of Technology, Tehran, Iran
mdhassanzd@aut.ac.ir
Mohammad Reza Meybodi
Computer Engineering and Information Technology Department, Amirkabir University of Technology, Tehran, Iran
mmeybodi@aut.ac.ir
Mohammad Mehdi Ebadzadeh
Computer Engineering and Information Technology Department, Amirkabir University of Technology, Tehran, Iran
ebadzadeh@aut.ac.ir

Abstract

This paper presents a modification of Particle Swarm Optimization (PSO) technique based on cooperative behavior of swarms and learning ability of an automaton. The approach is called Cooperative Particle Swarm Optimization based on Learning Automata (CPSOLA). The CPSOLA algorithm utilizes three layers of cooperation which are intra swarm, inter swarm and inter population. There are two active populations in CPSOLA. In the primary population, the particles are placed in all swarms and each swarm consists of multiple dimensions of search space. Also there is a secondary population in CPSOLA which is used the conventional PSO's evolution schema. In the upper layer of cooperation, the embedded Learning Automaton (LA) is responsible for deciding whether to cooperate between these two populations or not. Experiments are organized on five benchmark functions and results show notable performance and robustness of CPSOLA, cooperative behavior of swarms and successful adaptive control of populations.

Keywords: Particle Swarm Optimizer (PSO), Cooperative Particle Swarm Optimizer (CPSO), Learning Automata.

1. Introduction

Particle Swarm Optimization (PSO) [1], [2] is a population based technique inspired form shoaling behavior of fish and swarming behavior of insects. The mystery becomes evident when the simple rules that followed by individuals leads to emergent of well-organized system. Cooperative PSO (CPSO) [3], [4] is a variation of the traditional PSO algorithm in which the dimensions of population divided into multiple separate swarms and each swarm try to optimize the problem separately. During the fitness evaluation of particles, the cooperation is occurred between swarms. Comprehensive Learning PSO (CLPSO) [5] is one of the most successful PSO improvements. A new learning strategy is used in CLPSO, where all particles' best information is used to update any other particle's velocity. The inertia weight [6] is one of the most important PSO's parameters, which is used to balance the global and local search of the population. Recently, an Adaptive PSO (APSO) [7] has introduced. APSO adaptively controls the PSO parameters by estimating the population distribution. Beside the adaptation of the inertia weight, APSO algorithm controls acceleration coefficients by four strategies named as exploration, exploitation, convergence and jumping out.

A Learning Automaton (LA) [8], [9] is a machine which is adapted to changes in its environment. The adaption is the result of learning process of the automaton. Recently learning automata is used for adaptive parameter selection in Evolutionary Algorithms (EA) [10], [11]. Also a new hybrid method of optimization which called PSO-LA [12]–[17] has been emerged. In PSO-LA algorithms an LA or a group of learning automata is assigned to the whole population or each particle of the population. LA or group of LAs controls the path and velocity of the particles. Moreover, LA has application in Grid computing [18]. In [19], Distributed Learning Automata (DLA) has been used for Grid resource discovery.

CPSO family [3] consists of four algorithms: CPSO-S, CPSO-S_K, CPSO-H and CPSO-H_K where K is the split factor parameter which specifies the length of desired solution vector. Typically, while optimizing an N – dimensional problem by using CPSO-S, K will be set to N (number of dimensions). Having both beneficial characteristics of PSO and CPSO-S_K, CPSO-H_K is emerged as the combination of these two algorithms. It is a tempting idea to have a mechanism which is able to understand when to switch between PSO and CPSO-S_K [3].

In [16] the first attempt to improve this hybridization of PSO and CPSO algorithms is done by embedding an automaton as a toolbox of the switching mechanism. Furthermore, in this paper we deeply investigate the behavior of discussed learning automata approach for CPSO family by a set of diverse experiments.

* Corresponding Author

The paper is organized as follow: section 2 reviews two PSO heuristics. Section 3 introduces learning automaton and its application in PSO. Section 4 describes cooperative PSO based on learning automata. Experimental setup and simulation results are presented in section 5.

2. Particle Swarm Optimizer (PSO)

2.1 Conventional formulation of PSO

Particle Swarm Optimization (PSO) [1], [2] consists of a population of particles in which each particle represents a feasible solution vector. Assume that we have an N–dimensional problem space with M particles which are initialized in a feasible search space. The velocity, position and best previous position of i_{th} particle are respectively shown by $X_i = (x_i^1, x_i^2, \ldots, x_i^N)$, $V_i = (v_i^1, v_i^2, \ldots, v_i^N)$ and $pbest_i = (pbest_i^1, pbest_i^2, \ldots, pbest_i^N)$. Also the best position of the population is $gbest = (gbest^1, gbest^2, \ldots, gbest^N)$. The velocity V_i^d and position X_i^d of the d_{th} dimension of the i_{th} particle are manipulated through the following [5]:

$$V_i^d = w \times V_i^d + c_1 \times rand1_i^d \times \left(pbest_i^d - X_i^d \right) + c_2 \times rand2_i^d \times \left(gbest^d - X_i^d \right) \quad (1)$$

$$X_i^d = X_i^d + V_i^d \quad (2)$$

Where c_1 and c_2 are *acceleration constants* which absorb the particles to *pbest* and *gbest* positions. $w \in [0,1]$ is *inertia weight* which controls the global and local searches. *rand1* and $rand2 \in [0,1]$ are two random numbers generated for each dimension of the particles. The algorithm of the original PSO is given in Fig. 1.

2.2 Cooperative Learning in PSO

The idea of cooperative learning was first implemented in the field of Genetic Algorithm (GA) by Potter [20]. Potter suggested that for optimizing the designated target function, each dimension of the fitness function could be optimized by a distinct population and be evaluated in form of an N–dimensional vector through the fitness function. PSO and GA both suffer from the *Curse of dimensionality*. Using cooperative technique in PSO may lead to promising results. Recently The concept of cooperation mapped into PSO technique. Cooperative behavior in PSO was first introduced by Van den Bergh [4]. In cooperative PSO instead of having one swarm of M particles trying to optimize the designated N–dimensional optimization problem, we have N swarms of M particles which working on an isolated 1–dimensional problem. In this approach we should use a *context vector* to build a required N–dimensional vector to evaluate each of the swarms.

The family of CPSO algorithm proposed in [3] consists of the following algorithms: CPSO-S, CPSO-S_K, CPSO-H and CPSO-H_K. In CPSO-S algorithm each dimension of search space is considered as a swarm of M particles and all swarms are trying to find a better solution vector. If there is any correlation in the population, it would be desirable to gather the correlated dimensions in the same swarm. The idea of correlated variables leads to emergence of *split factor* parameter which tuned the swarm size. Now, instead of splitting the population into N swarms of 1-dimensional vectors like CPSO-S, we could have K swarms of C–dimensional vectors *(C<N)* like CPSO-S_K. Standard PSO algorithm has the ability of escaping from local minima and CPSO-S_K algorithm has fast convergence speed. Merging both beneficial characteristic of this two algorithms leads to appearance of CPSO-H_K algorithm. CPSO-H_K algorithm consists of two phases, in 1st phase CPSO-S_K run and the information exchange performs *from* CPSO-S_K half *to* PSO half of algorithm. At 2nd phase, PSO run and information exchange *form* PSO half *to* CPSO-S_K half performs. Note that each phase performs in a separate iteration.

Cooperative PSO [3], [4] divides the initial population into some subpopulations and each of these subswarms optimizes their designated dimensions individually. There are two layers of cooperation in a cooperative PSO. The first layer lies under the collaborative behavior of particles in specific dimensions and the second one is the schema that produces a solution vector by means of sharing the best information of each subpopulation to constitute a valid solution vector. In order to evaluate each member of the subpopulation, one requires constructing a *context vector* which aggregates the best solution of each subpopulation within an N-dimensional vector. Typically to evaluate the current subpopulation, the corresponding dimensions filled with the position of particle and the other dimensions are considered constant. Fig. 2 is the cooperative PSO pseudocode.

Algorithm 1   Standard PSO
for each generation k **do**
for each particle i **do**
Update velocity of i_{th} particle by (1)
Update position of i_{th} particle by (2)
Calculate particle fitness $f(x_i)$
Update $pbest_i$ and $gbest$
$i = i+1$ // next particle
end for
$k = k+1$ // next generation
end for

Fig 1. The pseudocode of the standard PSO

Algorithm 2   Cooperative PSO (CPSO)
define
Split N-dimensional search space into j subpopulations.
Calculate the best individual of each subpopulation (*sbest*).
Construct a *Context Vector* (**CV**) through the best individuals of each subpopulation:
CV = [*sbest₁*, *sbest₂*, …, *sbestⱼ*]
for each generation k **do**
for each subpopulation j **do**
for each particle i **do**
Replace current particle of jth Subpopulation by its

```
        corresponding positions in the CV
        Evaluate the N-dimensional output vector through
        the fitness function.
            i = i+1 // next particle
        end for
        Update sbestⱼ.
        j = j+1 // next swarm
      end for
      k = k+1 // next generation
    end for
```

Fig. 2 The pseudocode of the cooperative PSO

2.3 Cooperative based PSO algorithms

As well as cooperative PSO [21] and GA [22], the cooperative approach is also implemented in other EAs such as: Evolutionary Strategy (ES) [23], Differential Evolution (DE) [24], [25] and Artificial Bee Colony (ABC) [26]. The following are four recent advances in the context of cooperative PSO:

The Cooperative Coevolutionary ES (CCES) [23] divides the population of ES into some subspecies and lets them evolve. By means of a migration operator, CCES could hybridize the cooperative evolutionary behavior of CPSO with ES. The proposed model controls the interaction of subspecies properly and exhibits good performance results.

The Cooperative Coevolutionary DE (CCDE) [24] partitioned the problem into several sub problems and allocated a subpopulation to each of them. In [25], a randomized grouping mechanism introduced and an adaptive weighting strategy used in order to adapt the separated components. The idea was accomplished to bring the interacted variables into a similar subcomponent.

The self-adaptive neighborhood search into DE (SaNSDE) could tackle the non-separable problems with more than 1000 dimensions inside.

The cooperative approach of Potter is exerted into ABC and Cooperative ABC (CABC) [26] is emerged. Like two variants of CPSO, he introduced two versions of split swarm and hybrid for CABC. The CABC_S algorithm can efficiently optimize the separable problems and the CABC_H algorithm has the ability to escape from the local minima.

2.4 Evaluation Scheme of PSO versus CPSO

The key characteristic of Standard PSO [2] and cooperative PSO [3] is related to their corresponding population. The standard PSO contains a single population where this single population is divided into multiple swarms in cooperative PSO.

There is a paradigm in conventional PSO algorithm which could be extended to Cooperative PSO: *In order to find a proper solution vector, each particle of the swarm fly through an N–dimensional search space by N values corresponded to each dimension of the space*. To understand this phrase deeply, consider the population as a matrix $_{[M \times N]}$ where M and N respectively represent the number of particles and dimensions, respectively as: [*#of Particles* $\times$ *#of Dimensions*] (see Fig. 3). In this

framework, velocity and position of the standard PSO [2] population were updated row wise. The interpretation of this framework in CPSO [3] is quite different from that of standard PSO. In CPSO the population is optimized column wise (dimension wise) with the dimension of each particle being evaluated by a *context vector (CV)* which is built from the best particle of corresponding swarm and the best particles of other swarms.

$$
\begin{bmatrix}
P/D & \boxed{D_1} & \boxed{D_2} & \cdots & \boxed{D_N} \\
\boxed{P_1} & P[1,1] & P[1,2] & \cdots & P[1,N] \\
\boxed{P_2} & P[2,1] & P[2,2] & \cdots & P[2,N] \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\boxed{P_M} & P[M,1] & P[M,2] & \cdots & P[M,N]
\end{bmatrix}
$$

$$PSO: \quad f(P_i) = fitness(P_i(1 \rightarrow D))$$

$$CPSO: \quad f(P_i, S_j) = fitness(CV(P_i, S_j, j))$$

Fig. 3 Comprehensive view of the PSO population. $f(P_i)$ represents the i_{th} particle of population which evaluates through the traditional PSO mechanism and $f(P_i, S_j)$ indicates the evaluation process of the i_{th} particle (P_i) of j_{th} swarm (S_j) of CPSO population.

3. Learning Automata (LA)

3.1 Conventional Formulation of LA

Learning Automata [8], [9] is a stochastic optimization technique from the family of Reinforcement Learning (RL) algorithms. Having enough interaction with the unknown environment, elegance emerges and the optimal policy will be chosen. Fig. 4 shows how automaton interacts with its environment. A study of the learning process of LA in a random environment is comprehensively reported in [8].
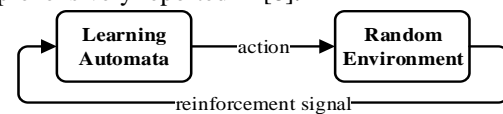
Fig. 4. The interaction between learning automata and environment.

Learning automata [8], [9] are divided into two groups of *fixed-structure* and *variable-structure automata*. A Variable-Structure LA (VSLA) is represented by a quadruple $[\alpha, \beta, p, T]$, Where $\alpha = \{\alpha_1, ..., \alpha_r\}$ is a set of actions, $\beta = \{\beta_1, ..., \beta_r\}$ is a set of inputs, $p = \{p_1, ..., p_r\}$ is the probability vector corresponds to each action and $p(n+1) = T[\alpha(n), \beta(n), p(n)]$ is the learning algorithm. If $p(n+1)$ is a linear function of $p(n)$ then the reinforcement scheme should be linear; otherwise it is nonlinear. In the simplest form of VSLA consider an automaton with r actions in a stationary environment where $\beta = \{0,1\}$ is included in inputs. After selecting the action by the automaton, the reinforcement signal will receive from the environment. When the *positive*

response ($\beta = 0$) received, the action probabilities are updated through (3):

$$p_j(n+1) = \begin{cases} p_j(n) + a.(1 - p_j(n)) & if\ i = j \\ p_j(n).(1-a) & if\ i \neq j \end{cases} \quad (3)$$

When the *negative response* ($\beta = 1$) received from the environment, action probabilities are updated through (4):

$$p_j(n+1) = \begin{cases} p_j(n).(1-b) & if\ i = j \\ \dfrac{b}{r-1} + (1-b).p_j(n) & if\ i \neq j \end{cases} \quad (4)$$

The a and b are called *learning parameters* and they are associated with the reward and penalty responses. If a and b are equal, the learning scheme is called L_{R-P} (Linear Reward-Penalty). If the learning parameter b is set to 0, then the learning scheme is named L_{R-I} (Linear Reward-Inaction). And finally if the learning parameter b is much smaller than a, the learning scheme is called $L_{R\varepsilon P}$ (Linear Reward-epsilon-Penalty).

3.2 LA based PSO algorithms

Parameter adaption [10], [11] is one of the most difficult tasks in EAs. As there are multiple parameters in PSO, it needs a mechanism to tune them during the evaluation of the population. In [10] a study of adaptive PSO parameter selection is conducted. Also, embedding learning automata in the population of PSO is another improvement; the model is called PSO-LA. In PSO-LA model an automaton is used to configure the search behavior of particles and adjust the velocity and position of them based on optimal selected policy. In coarse-grained PSO-LA [11] algorithms, an LA takes the responsibility of steering the whole swarm (|LA|=1). Since coarse-grained PSO-LA algorithms are trapped into local minima, in fine-grained PSO-LA algorithms [13], [14], learning automata are assigned to each particle of the swarm (|LA|=|*population size*|). This technique gives more maneuverability to the particles for searching through the search space. Some improved PSOs using LA are illustrated in the following:

A Dynamic Particle Swarm Optimization based on a 3-action Learning Automata (DPSOLA) introduced in [14]. The embedded learning automaton accumulates the information from individuals, local best and global best particles then combines them to navigate the particle through the problem space.

One variant of PSO is Comprehensive Learning Particle Swarm Optimizer (CLPSO) [5], which uses all individuals' best information to update their velocity. The novel strategy of CLPSO enables population to read from exemplars for specified generations which is called refreshing gap m. In [15], two classes of Learning Automata (LA) developed in order to study the learning ability of automata for CLPSO refreshing gap tuning. In the first class, a learning automaton is assigned to the population and in the second one each particle has its own personal automaton.

The cooperative PSO based on LA which introduced in [16] is the first version of CPSOLA. This algorithm utilizes both beneficial characteristics of PSO and CPSO by employing a learning automaton as a realtime decision making optimization tool.

The Adaptive Cooperative Particle Swarm Optimizer (ACPSO) [17] which facilitates cooperation technique through usage of LA algorithm. Cooperative learning strategy of ACPSO optimizes the problem collaboratively and evaluates it in different contexts. In ACPSO algorithm, a set of learning automata associated with dimensions of the problem are trying to find the correlated variables of the search space and optimize the problem intelligently. This collective behavior of ACPSO will fulfill the task of adaptive selection of swarm members.

4. Cooperative Particle Swarm Optimizer based on Learning Automata (CPSOLA)

4.1 Presenting Non-constraint Optimization Problems

Optimization is an approach to solve the complicated problems, like scheduling tasks. The framework of an N-dimensional, non-constraint optimization problem is as follows [5]:

$$\min\{f(x)\}; \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad x_1 \leq x \leq x_n \quad (5)$$

The optimization aim is to seek the optima in a search space by generating several feasible solutions and selecting the optimum one. Usually traditional techniques like dynamic programming and greedy algorithms are applied to optimization problems. In dynamic programming by using divide and conquer method, in each interval a portion of the search space will be omitted and the optimal solution will extract from the residual information. Also, during some steps of greedy algorithms, the most optimal solution will be selected by an excessive policy. In both of these listed algorithms, the election of each step is performed the local search with the hope of discovering the best solution. The lake of retaining the exploration and exploitation equilibrium and also excreting a segment of the search space in each epoch, are the defects of the conventional optimization approaches.
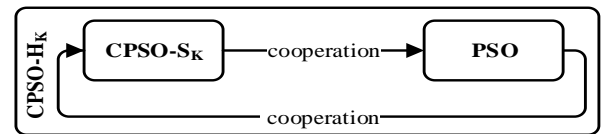

Fig. 5. Structural view of the CPSO-H_K algorithm.

4.2 Introducing the Non-adaptive Cooperative Scenarios

The CPSO model consists of four algorithms, from now on, our study specifically focused on CPSO-H_K algorithm which covers the other three ones. Fig. 5 shows the structure of CPSO-H_K algorithm. "When to switch between CPSO-S_K and PSO?" is a question proposed by Van den Bergh in [3]. For designing such a robust, general and adaptive mechanism consider the following scenarios in CPSO-H_K algorithm:

Scenario 1: At the earliest generations, the particles are scattered in the search space. It is desired to have fast global search of CPSO-S_K while any local minima are placed around the particles. At the middle iterations of the algorithm, while getting trapped in a local minimum, using PSO is much beneficial to escape from it.

Scenario 2: Immediately after escaping from local minima, it is safe for the algorithm to make use the high speed of CPSO-S_K till reaching the next local minimum.

Scenario 3: Sometimes the information exchange that is occurred in each generation of CPSO-H_K algorithm is unnecessary. This unnecessary amount of cooperation defects the run time performance of the CPSO-H_K.

Reviewing the discussed scenarios, interleave execution of CPSO-S_K and PSO seems to be a naïve form of cooperation between these two algorithms. A proper choice is to form an adaptive cooperation between CPSO-S_K and PSO algorithms. By using one learning automaton, we could have an adaptive switching mechanism which intelligently switches between CPSO-S_K and PSO algorithms. As well as preserving the positive characteristics of CPSO-S_K and PSO algorithms, CPSOLA algorithm significantly reduce the amount of information exchange.

4.3 Describing the Adaptive Cooperative Behavior of CPSOLA

Like CPSO-H_K in CPSOLA, we have two separate populations. The CPSO-S_K population is our primary population and the PSO population is the secondary one. Information exchange between two populations is postponed to *critical generations* because there is an adaptive switching mechanism between these two algorithms. A critical generation is a part of evolution process in which the cooperation between CPSO-S_K algorithm and PSO is vital. The CPSOLA algorithm presents two advantages in contrast to CPSO-H_k:

1) CPSOLA simultaneously utilizes both beneficial characteristics of PSO and CPSO-S_k. It inherits fast convergence speed of CPSO and also easily escapes from local minima by utilizing PSO's *gbest* and *pbest* information.

2) CPSOLA adaptively uses both CPSO-S_K and PSO in order to maintain marginal performance while CPSO-H_K interleave this algorithms without any environment perception.

3) CPSOLA balance the global and local search by maintaining population diversity between PSO and CPSO.

The scheme of adaptive switching mechanism of CPSOLA is shown in Fig. 6. The automaton has two actions: 1) Cooperation between primary and secondary population. 2) Isolation and just using primary population. In other words: 1) Running CPSO-H_K algorithm. 2) Running CPSO-S_K algorithm.
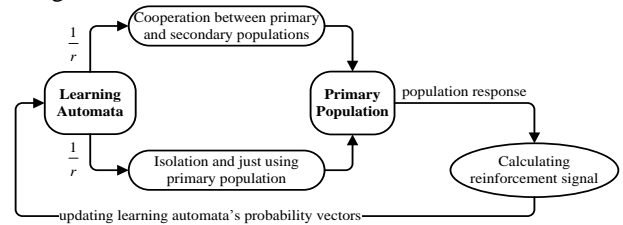
Fig. 6. Schematic view of decision making by learning automata

In each generation of CPSOLA algorithm, the automaton decides whether to start the alternative population (PSO population) or not. While having enough interactions with the populations, the automaton perceives when to switch between two algorithms. In the following we listed the five main differences between the CPSOLA model and CPSO model: 1) Instead of having unnecessary information exchange in each generation, preserve it for *critical generations*. 2) Reduced the workload of algorithm and made the execution time faster. 3) Preserve CPSO-S_K fast convergence speed property. 4) Keep the PSO's ability to escape from local minima. 5) Preserve the diversity of population.

Equation (6) is the criterion to evaluate the reinforcement signal in CPSOLA algorithm. *If* in the current iteration global best position of primary population (CPSO-S_K swarm best particle) improved *then* the automata's selected action would get the award and the automaton will be punished *otherwise*. Since the reinforcement signal is calculated in the context of primary population, the global best position of secondary population (PSO global best particle) do not have a direct influence on evaluating this signal.

$$\beta = \begin{cases} 0 & if \ fitness\left(Sbest_i < Sbest_{i-1}\right) \\ 1 & Otherwise \end{cases} \quad (6)$$

Fig. 7 is the pseudocode of CPSOLA algorithm. LA has two actions: first, cooperation or information exchange between CPSO and PSO population and second, isolation or evolution of CPSO population. In CPSOLA algorithm, an LA determines the time to perform cooperation between primary and secondary populations. At the earliest generations the LA selects actions randomly but after passing middle iterations the ultimate perception of the optimization problem occurs and the LA could accurately detects the time for switching between these two populations. This switching time is a key advantage of CPSOLA algorithm in contrast to the CPSO-H_K interleave switching strategy.

```
Algorithm 3 Cooperative PSO based on LA (CPSOLA)
define
   Initialize primary population with K swarms: P
   Initialize secondary population: Q
   Initialize LA with 2 actions: {cooperation, isolation}
do
   Select an action
   if the selected action is cooperation then
      for each swarm P_j; j ∈ [1..K]
         for each particle i ∈ [1..s]
         Update particle position by Equations (1) & (2)
            Calculate particle fitness
            Update pbest & gbest
            i = i + 1 // next particle
         end for
         j = j + 1 // next swarm
      end for
      Select a random particle form Q to write
      for each particle Q; i ∈ [1..s]
         Update particle position by Equations (1) & (2)
         Calculate particle fitness
         Update pbest & gbest
         i = i + 1 // next particle
      end for
      for each swarm P_j; j ∈ [1..K]
         Select a random particle form P to write
      end for
   else if the selected action is isolation then
      for each swarm P_j; j ∈ [1..K]
         for each particle i ∈ [1..s]
          Update particle position by Equations (1) & (2)
            Calculate particle fitness
            Update pbest & gbest
            i = i + 1 // next particle
         end for
         j = j + 1 // next swarm
      end for
   Evaluate reinforcement signal by Equation (6)
   Update LA's probability vectors by Equations (3) & (4)
k = k+1 // next generation
until a terminate condition is met
```

Fig. 7. The pseudocode of the CPSOLA algorithm.

4.4 Analyzing the Adaptive Cooperative Behavior of CPSOLA

Action selection of LA needed a comprehensive perception of the environment. In this section, a 30–dimensional Rosenbrock test function with 20 particles is used to investigate the interaction between LA and population during the evolution process. The fitness comparison of CPSO-S_K and PSO are plotted in Fig. 8.a. By looking on zoomed boxes of specific parts of evolution, we can observe that, in the earliest iterations of CPSOLA algorithm, there is no obvious difference between the fitness of two populations. The algorithm probably could escape from the local minima after reaching the middle iterations, so the policy is changed adaptively and even during this part of evolution the PSO's fitness even could be better than CPSO-S_K's fitness. Because of its cooperative search method of CPSOLA, in the last generations, the CPSO-S_K's fitness becomes

dominated and the algorithm converges faster than original CPSO algorithm.

The variance of action probabilities are plotted in Fig. 8.b. *Isolation* action means the algorithm is just used its primary population, hence the *Cooperation* action means two populations perform information exchange. It can be seen that CPSOLA algorithm has the ability to jump out of the local optima, which is the result of hidden diversity of PSO algorithm. Although the first and the second population perform cooperation but, while the learning automaton selects the *isolation* action, the PSO population skips some of the iterations. Since the cooperation dose not performs during each iteration, this trend seems to be a little unwanted. But by looking from the outer layer of cooperation, writing a bad fitness from alternative population into primary one could increase the diversity of primary population. In the middle generations, while primary population is stagnated in a local minimum, CPSOLA algorithm starts exchanging the information between two populations. This means that in a few generations, the secondary population could overwrite its inferior solutions to the primary population except the global best particle of each swarm which is protected. The diversity of primary population will increase significantly and the algorithm easily could escape from the local minimum.
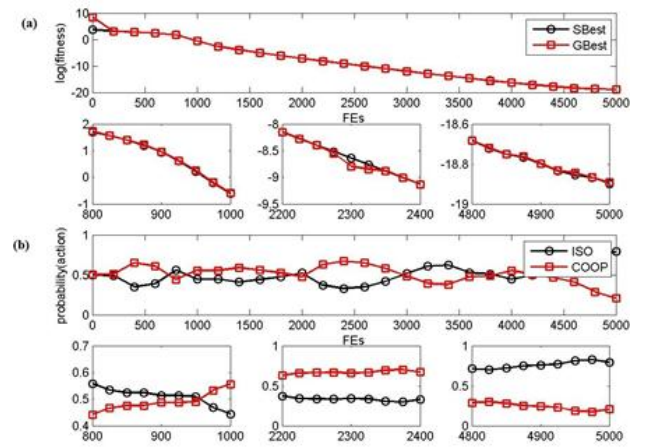
Fig. 8. (a) The CPSO-S_K's Swarm Best Position (*sbest*) versus the PSO's global best position (*gbest*). The horizontal and vertical axis of each box represents the number of fitness evaluations (FEs) and the fitness value in logarithmic scale, respectively. (b) Variance of LA's actions probability. The horizontal and vertical axis of each box represents the number of fitness evaluations (FEs) and the probability value in Gaussian scale, respectively.

5. Experimental Study

5.1 Simulation Setup

In order to have a fair comparison, all the PSOs should use a same number of fitness. This number is set to 20000. All experiments were run 10 times; the means and variances of best solution of these runs are reported. In order to study the impact of population size; the experiments repeated with 10, 15 and 20 particles per

population. All benchmark functions are 30–dimensional optimization problems.

Observing the learning automaton's [10], [11] behavior during the evolution process, three different kind of learning algorithms are placed in CPSOLA algorithm [16]. The detailed configurations of the reward and penalty parameters are mentioned in Table I. As long as the L_{RP} learning algorithm acts as a moderate one among three of them, in the following we just report the results of this learning algorithm.

In order to compare the proposed method, we simulate four other PSOs including: standard PSO algorithm [2], Split swarm Cooperative PSO (CPSO-S) [3], Hybrid Cooperative PSO (CPSO-H) [3] and Comprehensive Learning PSO (CLPSO) [5]. The settings of these PSOs are briefly listed in Table II.

We choose five benchmark functions [3], [5] to run the experimental tests. Among them, f_0 (Rosenbrock) and f_1 (Quadric) functions are simple unimodal problems which have non-complex structure. The functions f_2 (Ackley), f_3 (Rastrigin) and f_4 (Griewank) functions are highly multimodal problems with many local optima positioned in their grid. The details and mathematical formulation of these benchmark functions are depicted in Table III.

Unimodal functions have simple structure and small number of minima. These kinds of optimization benchmarks can be easily optimized by deterministic optimization algorithms which use gradient information of benchmark. The first problem, function f_0 is a non-convex function that its global optima is placed in a long, narrow and hyperbolic valley. This function can be treated as a multimodal function. The function f_1 is a simple function that usually used for showing the power of optimization algorithms in function optimization. It is a separable benchmark function that each dimension could optimize independently.

Table I: The parameters configuration of learning automata. The **LA** column indicates the learning algorithm. The **Alpha** and **Beta** columns indicate different values of reward and penalty signals for each learning algorithm. Also the **Action set** column shows two actions (*cooperation and isolation*) of learning automata and the **Initial probability** of 0.5 indicates the raw probability of each action

LA	Alpha	Beta	Action set	Initial probability
L_{RP}	0.01	0.01	{cooperation, isolation}	{0.5, 0.5}
$L_{R\varepsilon P}$	0.001	0.01	{cooperation, isolation}	{0.5, 0.5}
L_{RI}	0.01	0.00	{cooperation, isolation}	{0.5, 0.5}

Table II: The PSO algorithms used for simulation. Each algorithm has unique parameter settings which are derived from its associated reference. **The Algorithm** column lists the different PSOs that used for simulation. The **Parameters** column includes the parameter used for each PSO. The **Topology** column shows the special attribute of each PSO and the **Ref**. column represents the associated reference number

Algorithm	Parameters	Topology	Ref.
PSO	$w = 0.72, c_1=c_2=2.0$	Global version	[2]
CPSO-S_K	$w:0.9-0.4, c_1=c_2=1.49,K=6$	Cooperative swarms	[3]
CPSO-H_K	$w:0.9-0.4, c1=c2=1.49,K=6$	Hybrid Cooperative swarms	[3]
CLPSO	$w:0.9-0.4, c=1.49445,m=7$	Comprehensive Learning Strategy	[4]
CPSOLA	$w:0.9-0.4, c_1=c_2=1.49,K=6$	Adaptive Hybrid Cooperative swarms	[-]

Most of optimization heuristics suffer from the *curse of dimensionality* [27]. This phenomenon appears when the performance of algorithm degraded rapidly by increasing the dimensionality of the problem. This situation has two reasons: 1) By increasing the dimensionality of benchmark functions, the number of local minima increases exponentially. A successful algorithm in this situation is one that can search more promising regions of search space. 2) Some benchmark functions are reshaped by increasing the number of dimensions. For example, function f_2 is a unimodal function in low number of dimensions (2 dimensions). Also, in this function by increasing the number of dimensions it will convert to a multimodal function. Due to the aforementioned reasons of curse of dimensionality, some search strategies which may work well in low dimensional benchmark functions can't find the optimum solution in high dimensional benchmark functions.

Multimodal benchmark function (f_2-f_4) has many local optima and one global optimum. These benchmark functions have complicated structure. The convergence speed of multimodal functions is lower than unimodal functions.

The function f_2 is a separable function with an exponential term which covers the surface of function with many local minima. It has one narrow global optimum basin and many minor local optima. Simple gradient descent algorithms are failed to optimize this function, but any heuristic which can move through the valley of the function can attain better results.

The function f_3 is a complex multimodal problem with large number of local optima. When attempting to solve function f_3, algorithms may easily fall into a local optimum. Hence, an algorithm capable of maintaining larger population diversity is likely to yield better results.

The function f_4 has a $\prod_{i=1}^{n} cos\left(\frac{x_i}{\sqrt{i}}\right)$ component causing linkages among variables, thereby making it difficult to reach the global optimum. An interesting phenomenon of function f_4 is that it is more difficult for lower dimensions than higher dimensions.

Table III: The formulation of benchmark functions. The **Function** column shows five different benchmark functions. The **Formula** column shows the mathematical formulation of each benchmark function. The values listed in the **Range** column are used to specify the magnitude to which the initial random particles are scaled. The V_{max} column indicates the maximum velocity which is used to clamp the velocity of particles. The **Threshold** column lists the function value threshold which is used as a stopping criterion in section 5-4. The **Shape** column shows a 2D view of each benchmark function
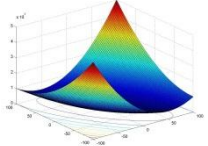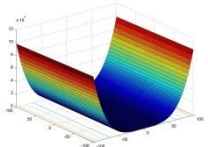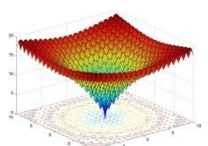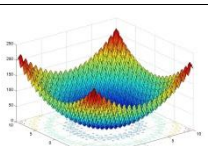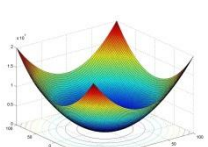
Function	Formula	Range	V_{max}	Threshold	Shape
f_0 (Rosenbrock)	$f_0(x) = \sum_{i=1}^{\frac{n}{2}} \left(100\left(x_{2i} - x_{2i-1}^2\right)^2 + \left(1 - x_{2i-1}\right)^2 \right)$	[-2.048, 2.048]	2.048	0.01	
f_1 (Quadric)	$f_1(x) = \sum_{i=1}^{n} \left(\sum_{j=1}^{i} x_j \right)^2$	[-100, 100]	100	0.01	
f_2 (Ackley)	$f_2(x) = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \right)$ $-\exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos\left(2\pi x_i\right) \right) + 20 + e$	[-32, 32]	32	0.01	
f_3 (Rastrigin)	$f_3(x) = \sum_{i=1}^{n} \left(x_i^2 - 10\cos\left(2\pi x_i\right) + 10 \right)$	[-5.12, 5.12]	5.12	0.00001	
f_4 (Griewank)	$f_4(x) = \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n}\cos\left(\frac{x_i}{\sqrt{i}} \right) + 1$	[-600, 600]	600	0.01	

Table IV: The average and standard deviation of PSO, CLPSO, CPSO-S$_6$, CPSO-H$_6$, CPSOL$_{RP}$, CPSOL$_{Re\ P}$ and CPSOL$_{RI}$ algorithms over 10 independent runs on 5 benchmark 30 – dimensional functions with 20000 Fitness Evaluations. The second column **S** lists the number of particles per swarm

Function Algorithm	S	f_0 (Rosenbrock)	f_1 (Quadric)	f_2 (Ackley)	f_3 (Rastrigin)	f_4 (Griewank)
PSO	10	1.30E-01 ± 1.45E-01	1.08E+00 ± 1.41E+00	7.33E+00 ± 6.23E-01	8.27E+01 ± 5.64E+00	9.65E-01 ± 7.58E-01
	15	5.53E-03 ± 6.19E-03	2.85E-72 ± 5.41E-72	4.92E+00 ± 5.81E-01	7.44E+01 ± 5.66E+00	2.62E-01 ± 1.61E-01
	20	9.65E-03 ± 7.28E-03	2.17E-98 ± 4.20E-98	3.57E+00 ± 4.58E-01	6.79E+01 ± 4.84E+00	6.51E-02 ± 2.17E-02
CLPSO	10	5.12E+00 ± 3.23E+00	2.96E+02 ± 1.78E+02	6.45E+00 ± 1.42E+00	1.74E+01 ± 4.60E+00	7.27E-01 ± 1.28E+00
	15	2.22E+00 ± 1.04E+00	9.79E+01 ± 6.98E+01	3.30E+00 ± 1.37E+00	7.26E+00 ± 2.85E+00	1.62E-02 ± 3.07E-02
	20	1.88E+00 ± 3.26E-01	4.43E+01 ± 1.33E+01	1.91E+00 ± 4.33E-01	3.68E+00 ± 2.10E+00	**5.64E-03 ± 1.40E-02**
CPSO-S$_6$	10	1.41E+00 ± 4.73E-01	4.63E-07 ± 6.14E-07	1.12E-06 ± 4.01E-07	0.00E+00 ± 0.00E+00	7.29E-02 ± 1.49E-02
	15	2.47E+00 ± 7.00E-01	1.36E-05 ± 1.76E-05	1.11E-05 ± 4.53E-06	0.00E+00 ± 0.00E+00	6.90E-02 ± 1.56E-02
	20	1.59E+00 ± 5.01E-01	1.20E-04 ± 8.99E-05	5.42E-05 ± 1.66E-05	0.00E+00 ± 0.00E+00	8.95E-02 ± 1.68E-02
CPSO-H$_6$	10	1.94E-01 ± 2.63E-01	2.63E-66 ± 5.08E-66	9.42E-11 ± 7.58E-11	0.00E+00 ± 0.00E+00	6.75E-02 ± 1.40E-02
	15	2.59E-01 ± 2.47E-01	9.00E-46 ± 1.09E-45	9.57E-12 ± 7.96E-12	0.00E+00 ± 0.00E+00	5.54E-02 ± 1.27E-02
	20	4.21E-01 ± 3.21E-01	1.40E-29 ± 1.15E-29	2.73E-12 ± 2.03E-12	0.00E+00 ± 0.00E+00	5.24E-02 ± 1.19E-02
CPSOL$_{RP}$	10	3.76E-23 ± 8.43E-23	5.09E-229 ± 0.00E+00	5.42E-14 ± 1.23E-14	0.00E+00 ± 0.00E+00	3.33E-02 ± 3.83E-02
	15	**1.33E-26 ± 4.78E-27**	3.07E-302 ± 0.00E+00	**4.99E-14 ± 7.64E-15**	**0.00E+00 ± 0.00E+00**	2.38E-02 ± 3.07E-02
	20	1.14E-26 ± 3.20E-27	**3.32e-321 ± 0.00E+00**	5.28E-14 ± 1.06E-14	0.00E+00 ± 0.00E+00	4.10E-02 ± 3.95E-02

5.2   Experiment 1: Function Optimization

Table IV shows the mean and standard deviation of 10 runs of each PSO on 30 dimensional problems. In unimodal functions cooperative PSOs could reach good results. The function f_0 is a simple unimodal function and its global optima places in the center of function. Also, function f_1 is a unimodal function with one global optimum. Standard PSO performs better in f_1 when comparing with CPSO family and CLPSO. This implies that the CPSO is less effective in solving simple problems. In the other hand, due to adaptive switching mechanism of CPOSLA, the algorithm could use the PSO's dimension-wise updating rule while using several swarms to optimize different dimensions of search space independently. By combining the results of these two properties, CPSOLA achieves the best results on unimodal functions (f_0 and f_1).

In Table IV, functions f_2-f_4 are multimodal functions. In f_2 and f_3 CPSOLA and CPSO generally outperform standard PSO, CLPSO that involves neither cooperative swarms nor information exchange property. However, the CPSOLA is most powerful and robust for these two test problems. These results confirm the hypothesis that cooperative swarms speed up the convergence of CPSOLA algorithm and adaptive switching mechanism helps the swarms jump out of the local optima and find better solutions.

Different dimensions of CLPSO may learn from different exemplars. Due to this, the CLPSO explores a larger search space than the other PSOs in function f_4. The larger search space is not achieved randomly.

Instead, it is based on the historical search experience. Because of this, the CLPSO performs comparably to or better than other PSO variants on function f_4. Note that function f_4 is known to become easier as the number of dimensions increases.

The experiments which are conducted on 30–D problems are repeated with 10, 15 and 20 particles. In Table IV, the S entry indicates the population size. The results show, increasing the number of particles and keeping the problem's dimension fixed, will lead to improve the fitness value of PSOs. Although from the results of Table IV the resulting performance may vary depending on the problem being optimized. There seems to be no definitive value for the swarm size that is optimal across all problems, so to avoid tuning the algorithm to each specific problem, a compromise must be reached. 15 particles were shown the best results for CPSOLA, as populations of this size performed best by a very slight margin when averaged across the entire range of test problems.

5.3   Experiment 2: Convergence graph

Fig. 9 presents the convergence characteristics in terms of the mean best fitness value of each algorithm for all 30 – dimensional test function with 20 particles. The comparisons in both Table IV and Fig. 9 show that, when solving unimodal and multimodal problems, the CPSOLA offers the best performance on most test functions. In particular, the CPSOLA offers the highest accuracy on functions f_0, f_1, f_2 and f_3. Furthermore, CLPSO shows the best convergence characteristics on function f_4.
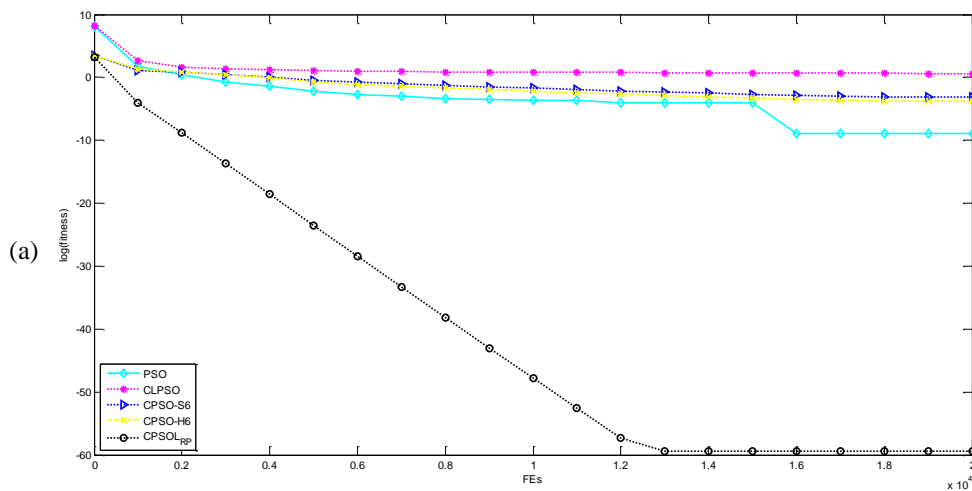

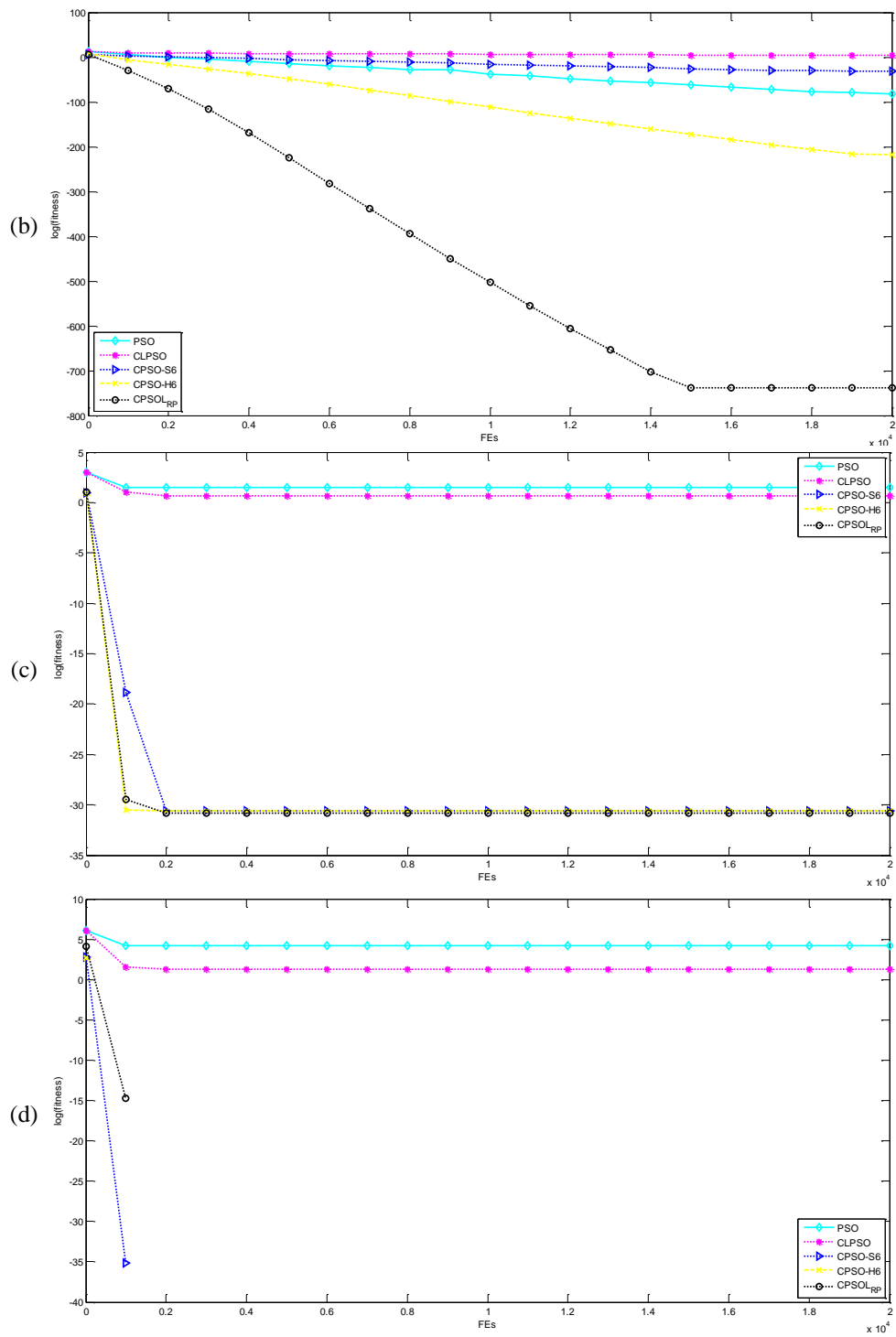Fig. 9. The mean convergence graph of 30-D benchmark functions. (a) f_0 (Rosenbrock).

Fig. 9. (Continued.) The mean convergence graph of 30-D benchmark functions. (b) f_1 (Quadric). (c) f_2 (Ackley). (d) f_3 (Rastrigin).
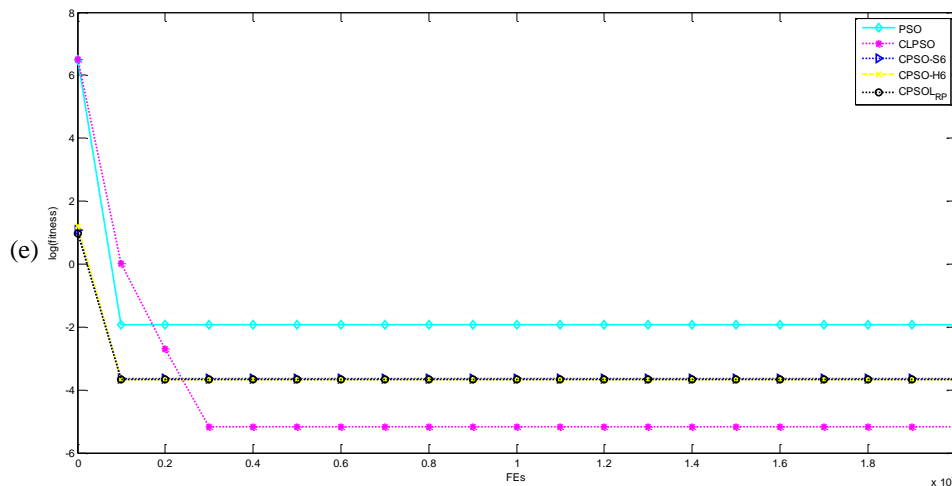
Fig. 9. (Continued.) The mean convergence graph of 30-D benchmark functions (e) f_4 (Griewank)

The Rosenbrock function (f_0) is a non-convex test problem with a long, narrow and parabolic shaped flat valley. From the set of flat lines of Fig. 9 – a (f_0), CLPSO and CPSO are easily get trapped in local minima while standard PSO still improves its fitness value. Moreover, CPSOLA could escape from the local minima till the middle generations and acquire the best fitness value. CPSOLA can find the best trail into the valley and almost converges to the global minimum.

The Quadric function (f_1) is a convex and unimodal problem which has D local minima except for the global one. In Fig. 9 – b (f_1) CPSOLA and CPSO-H_6 converge better than other PSOs. Both of these cooperative heuristics utilize two populations in order to optimize the test problem. Since CPSOLA uses adaptive switching mechanism while CPSO-H_6 uses interleave switching mechanism, the CPSOLA approach could exploit beneficial properties of PSO and CPSO-S_k algorithms.

The Ackley function (f_2) is a non-separable and multimodal function. It has a flat outer region and a large hole at the center. The function poses a risk for optimization heuristics, particularly hill climbing and gradient steepest descent algorithms. Fig. 9 – c (f_2) shows the convergence characteristics of 30– dimensional Ackley' function with 20 particles. The first flat line in Fig. 9 – c indicates that the Standard PSO becomes trapped in a local minimum in early generations. Since CLPSO has a large feasible search space, it is easily trapped in a local minimum either; the second flat line shows that. From the third and fourth flat lines which belong to CPSO-S_6 and CPSO-H_6 algorithms, we can observe that they have a fast convergence speed and these results are due to the exhaustive dimension wise search method of cooperative approach. The adaptive switching mechanism of CPSOLA has a cost and the cost is the slow convergence of the algorithm in this problem. Since

using the alternative population may suppress improving the global best particle of primary population for a while, CPSOLA is managed to continue improving its performance very well. The best result belongs to CPSOL$_{RP}$, which its learning algorithm can find the optimal policy faster than the others. Also definitive decision making property of LRP learning algorithm helps CPSOL$_{RP}$ algorithm to escape form the local minima before it becomes too late.

The Rastrigin function (f_3) is a non-convex, non-linear and multimodal optimization problem that has large number of local minima whose values increases with the distance to the global minimum. Finding the minimum of Ackley function is fairly difficult problem due to Ackley's large search space and its large number of local minima. From Fig. 9 – d (f_3), PSO and CLPSO are get trapped in local minima in early generations but all cooperative PSOs converge to global optimum. Also the CPSOLA algorithm can converge faster than CPSO-S_6.

The Griewank function (f_4) is a multimodal and non-separable test function. The algorithms which try to optimize each variable of this benchmark function independently, lead to failure. The global optimum of this function is regularly distributed. From the results presented in Fig. 10 – e (f_4), it can be observed that all PSO variants failed on Griewank function except for CLPSO. The algorithm utilizes a tournament selection heuristic for determining the exemplar in each dimension.

Figures 9 – a to e show the plot performance of the PSO, CLPSO, CPSO-S_6, CPSO-H_6 and CPSOL$_{RP}$ algorithms in 30-dimensional problems. The CPSOLA algorithm uses a bi-action learning automaton with the L_{R-P} learning algorithm.

Table V: Robustness Analysis. The second column **S** lists the number of particles per swarm. The column **Succeeded** list the number of runs (out of 10) that manage to attain a function value below the threshold in less than 20000 fitness evaluations, while the column **FEs** presents the number of function evaluations needed on average to reach the threshold, calculated only for runs that **Succeeded**

Function Algorithm	S	f_0 (Rosenbrock)		f_1 (Quadric)		f_2 (Ackley)		f_3 (Rastrigin)		f_4 (Griewank)	
		Succeeded	FEs.	Succeeded	FEs.	Succeeded	FEs.	Succeeded	FEs.	Succeeded	FEs.
PSO	10	10	55	1	1325	1	24	5	100	1	85
	15	10	53	10	4209	4	77	8	128	4	409
	20	10	48	10	2779	5	96	9	182	5	262
CLPSO	10	0	N/A	0	N/A	0	N/A	0	N/A	0	N/A
	15	0	N/A	0	N/A	0	N/A	0	N/A	7	5240
	20	0	N/A	0	N/A	0	N/A	1	238	**8**	3391
CPSO-S_6	10	10	1377	10	555	10	145	10	303	4	319
	15	10	1306	10	450	10	118	10	253	2	138
	20	10	1185	10	398	10	102	10	211	2	116
CPSO-H_6	10	10	1067	10	248	10	61	10	103	2	40
	15	10	1043	10	214	10	56	10	99	2	45
	20	10	992	10	201	10	49	10	101	2	49
CPSOL$_{RP}$	10	10	1149	10	306	10	77	10	143	1	33
	15	10	1145	10	266	10	70	10	144	2	63
	20	**10**	1075	**10**	258	**10**	65	**10**	137	3	116

5.4 Experiment 3: Robustness Analysis

This section compares the various PSOs to determine their relative rankings using both *robustness* and convergence speed as criteria. The term robustness is used here to mean that the algorithm succeeded in reducing the function value below a specified threshold using fewer than the maximum allocated number of function evaluations. A *robust* algorithm is one that can decrease the fitness value below a specified threshold in a fewer number of fitness evaluations during all runs [3].

Table V shows the robustness analysis for all test functions. In function f_0 none of the algorithms, with the exception of CLPSO, have any difficulty reaching the threshold during any of the runs. Table V further shows that all the algorithm solved this problem in fewer than 2000 function evaluations, with the PSO algorithm requiring the fewest function evaluations overall.

The CLPSO algorithm again consistently fails to reach the threshold value of function f_1 on all runs. Furthermore, standard PSO with 10 particles per population is not a robust algorithm, whereas all cooperative algorithms reach the threshold during all runs.

The standard PSO and CLPSO have some difficulty with f_2 function, as can be seen in Table V. The function f_2 represents a very important result regarding the nature of cooperative algorithms: the CPSOLA algorithm may have somewhat slower rates of convergence compared with CPSO-S_6 and CPSO-H_6 algorithms, but it is significantly as robust as them in many cases. The function f_3 shows similar results to function f_2. The cooperative algorithms again perform admirably on the Rastrigin function, but the PSO and CLSPO algorithms are less robust in this test function.

The function f_4 shows interesting results. It proves to be hard to solve for all the algorithms, as can be seen in

Table V. None of cooperative algorithms and standard PSO can perfectly optimize it, while CLPSO's novel learning strategy enables it to act more robust when compared to other PSO variants. Also, no algorithm could achieve a perfect score on this test function.

When looking at the number of function evaluations, the CPSO-H_6 algorithm was usually the fastest, followed by the CPSOLA and the CPSO-S_6. These results indicate that there is a tradeoff between the convergence speed and the robustness of the algorithm.

In 4 out of 5 test functions, if the number of fitness evaluations reaches more than 5000 FEs, the algorithm will not meet the robustness condition, while in function f_4 this observation rejects and CLPSO still satisfy the robustness criteria.

5.5 Experiments Analysis

From Table IV, the CPSOLA performs better in 4 out of 5 benchmark functions, where swarm size were 15. This swarm size indicates that the algorithm does not need big population in order to gain better results compared with other PSO variants. Also the CPSOLA performs better while optimizing both unimodal and multimodal optimization benchmarks. Based on Table IV results the algorithm performs better while optimizing complex multimodal optimization benchmarks such as Ackley and Rastrigin. More over from Table V, the proposed algorithm is robust while optimizing complex multimodal functions where there optimization problems have complicated structures.

From the point of complexity analysis, the time complexity standard PSO [2] is $O(g * p * d)$ where n is the number of **g**enerations, m is the number of designated **p**articles and d is the number of **d**imensions. Totally, we assert that the time complexity of PSO is $O(n^3)$.

In this context by considering the Fig. 7, as long as we may have two choices in each generation the time complexity of CPSOLA will be calculated through the following:

$$O(g) * \begin{cases} O(s) * O(s_d) + O(p) \\ O(s) * O(s_d) \end{cases} \quad (7)$$

Where is (7), s is the number of swarms $(s<<p)$ and s_d is the number of member dimensions of sth swarm $(s_d<<d)$. In order to calculate the time complexity of CPSOLA, we consider the maximum of the two terms which is mentioned in (7), thus we have: $O(g * S * Sd + g * p)$. Totally we assert that the time of complexity of CPSOLA is $O(n^{3/S4}+n^2)$. As long as in each generation of CPSOLA algorithm the sub regions of optimization space is optimized through by swarms, the $O(n^3)$ value of PSO is divided by this sub swarms. So, in the worst case that the number of swarms is set to the whole population, the time complexity will be $O(n^3+n^2)$; where this value is worse than the time complexity of standard PSO. But also in the moderate case or best case where the solution space is divided into s_d swarms where $\lim_{s_d\to\infty}\frac{n}{s_d} = 0$ the algorithm may performs faster than standard PSO with time complexity of $O(n + n^2)$.

6. Conclusions

In this paper we presented a cooperative particle swarm optimizer with adaptive control on the outer layer of cooperation named as CPSOLA. The results are shown a significant improvement in performance and robustness. Like CPSO-H_K algorithm in CPSOLA we have two populations: The first one named as primary population and belongs to CPSO, while the secondary one belongs to PSO algorithm. In the proposed approach a learning automaton observe the global best fitness of primary population and decide when to cooperate with secondary one. Having a comprehensive scheme of problem to be optimized, the learning automaton controls the evolution process. Since the evolution of secondary population may lag from the first one, the automaton brings an indirect diversity while switching between its actions. In the real world every action has a consequence; slow convergence is the cost that we paid for our algorithm.

To evaluate the performance of CPSOLA, three different kinds of experiments are conducted in this paper. The experiments are carried out on five algorithms on the five chosen test problems belonging to two classes. The CPSOLA utilizes the L_{RP} learning technique as its learning algorithm. The CPSOLA performs the best in unimodal test functions and two out of three multimodal test functions. Totally the CPSOLA is shown the best performance in four out of five test functions. Also, it is significantly a robust algorithm with small standard deviation.

References

[1] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proceedings of the Sixth International Symposium on Micro Machine and Human Science, 1995. MHS '95*, 1995, pp. 39–43.

[2] D. Bratton and J. Kennedy, "Defining a Standard for Particle Swarm Optimization," in *IEEE Swarm Intelligence Symposium, 2007. SIS 2007*, 2007, pp. 120–127.

[3] F. van den Bergh and A. P. Engelbrecht, "A Cooperative approach to particle swarm optimization," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, pp. 225– 239, Jun. 2004.

[4] F. van den Bergh and A. P. Engelbrecht, "Cooperative learning in neural networks using particle swarm optimizers," *South African Computer Journal*, pp. 84–90, 2000.

[5] J. Liang, A. Qin, P. N. Suganthan, and S. Baskar, "Comprehensive learning particle swarm optimizer for global optimization of multimodal functions," *Evolutionary Computation, IEEE Transactions on*, vol. 10, no. 3, pp. 281–295, 2006.

[6] A. Nickabadi, M. M. Ebadzadeh, and R. Safabakhsh, "A novel particle swarm optimization algorithm with adaptive inertia weight," *Applied Soft Computing*, vol. 11, no. 4, pp. 3658–3670, Jun. 2011.

[7] Z. H. Zhan, J. Zhang, Y. Li, and H. S. H. Chung, "Adaptive particle swarm optimization," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 39, no. 6, pp. 1362–1381, 2009.

[8] K. S. Narendra and M. A. L. Thathachar, *Learning automata: an introduction*. Prentice-Hall, Inc., 1989.

[9] C. Ünsal, "Intelligent navigation of autonomous vehicles in an automated highway system: Learning methods and interacting vehicles approach," Virginia Polytechnic Institute and State University, 1997.

[10] A. B. Hashemi and M. R. Meybodi, "A note on the learning automata based algorithms for adaptive parameter selection in PSO," *Applied Soft Computing*, vol. 11, no. 1, pp. 689–705, Jan. 2011.

[11] A. Rezvanian and M. R. Meybodi, "LACAIS: Learning Automata based Cooperative Artificial Immune System for Function Optimization," in *3rd International Conference on Contemporary Computing (IC3 2010), Noida, India. Contemporary Computing*, CCIS, 2010, vol. 94, pp. 64–75.

[12] M. Sheybani and M. R. Meybodi, "PSO-LA: A New Model for Optimization," in *Proceedings of 12th Annual CSI Computer Conference of Iran*, 2007, pp. 1162–1169.

[13] M. Hamidi and M. R. Meybodi, "New Learning Automata based Particle Swarm Optimization Algorithms," presented at the Iran Data Mining Conference (IDMC), 2008, pp. 1–15.

[14] M. Hasanzadeh, M. R. Meybodi, and S. Shiry, "Improving Learning Automata based Particle Swarm: An Optimization Algorithm," in *12th IEEE International Symposium on Computational Intelligence and Informatics*, Budapest, 2011.

[15] M. Hasanzadeh, M. R. Meybodi, and M. M. Ebadzadeh, "Adaptive Parameter Selection in Comprehensive Learning Particle Swarm Optimizer," presented at the Symposium on Artificial Intelligence and Signal Processing (AISP), Tehran, Iran, 2013, pp. 1–10.

[16] M. Hasanzadeh, M. R. Meybodi, and M. M. Ebadzadeh, "A robust heuristic algorithm for Cooperative Particle Swarm Optimizer: A Learning Automata approach," in *2012 20th Iranian Conference on Electrical Engineering (ICEE)*, 2012, pp. 656 –661.

[17] M. Hasanzadeh, M. R. Meybodi, and M. M. Ebadzadeh, "Adaptive cooperative particle swarm optimizer," *Appl Intell*, vol. 39, no. 2, pp. 397–420, Sep. 2013.

[18] M. Hasanzadeh and M. R. Meybodi, "Deployment of gLite middleware: An E-Science grid infrastructure," in *2013 21st Iranian Conference on Electrical Engineering (ICEE)*, 2013, pp. 1–6.

[19] M. Hasanzadeh and M. R. Meybodi, "Grid resource discovery based on distributed learning automata," *Computing*, pp. 1–14.

[20] M. Potter and K. De Jong, "A cooperative coevolutionary approach to function optimization," *Parallel Problem Solving from Nature—PPSN III*, pp. 249–257, 1994.

[21] J. Kennedy and R. Eberhart, "Particle swarm optimization," in , *IEEE International Conference on Neural Networks, 1995. Proceedings*, 1995, vol. 4, pp. 1942–1948.

[22] J. H. Holland, "Genetic algorithms," *Scientific American*, vol. 267, no. 1, pp. 66–72, 1992.

[23] T. Bäck and H. P. Schwefel, "An overview of evolutionary algorithms for parameter optimization," *Evolutionary computation*, vol. 1, no. 1, pp. 1–23, 1993.

[24] M. F. Han, S. H. Liao, J. Y. Chang, and C. T. Lin, "Dynamic group-based differential evolution using a self-adaptive strategy for global optimization problems," *Applied Intelligence*, pp. 1–16, 2012.

[25] Z. Yang, K. Tang, and X. Yao, "Large scale evolutionary optimization using cooperative coevolution," *Information Sciences*, vol. 178, no. 15, pp. 2985–2999, 2008.

[26] M. El-Abd, "A cooperative approach to The Artificial Bee Colony algorithm," in *2010 IEEE Congress on Evolutionary Computation (CEC)*, 2010, pp. 1–5.

[27] R. Bellman, "Dynamic programming and Lagrange multipliers," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 42, no. 10, p. 767, 1956.

Mohammad Hasanzadeh received the B.Sc degree in Software Engineering from South Khorasan Payame Noor University (SKPNU), Birjand, Iran, in 2009 and M.Sc in Artificial Intelligence from Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran in 2013. His current research interests include computational intelligence, Machine Learning and Grid Computing.

Mohammad Reza Meybodi received the B.Sc and M.Sc degrees in Economics from Shahid Beheshti University, Tehran, Iran, in 1973 and 1977, respectively. He also received the M.Sc and Ph.D. degrees from the Oklahoma University, USA, in 1980 and 1983, respectively, in Computer Science. Currently he is a Full Professor in Computer Engineering Department, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. Prior to current position, he worked from 1983 to 1985 as an Assistant Professor at the Western Michigan University, and from 1985 to 1991 as an Associate Professor at the Ohio University, USA. His research interests include channel management in cellular networks, learning systems, parallel algorithms, soft computing and software development.

Mohammad Mehdi Ebadzadeh received the B.Sc in Electrical Engineering from Sharif University of Technology, Tehran, Iran in 1991 and M.Sc in Machine Intelligence and Robotic from Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran in 1995 and his Ph.D. in Machine Intelligence and Robotic from Télécom ParisTech, Paris, France in 2004. Currently, he is an Associate Professor in Computer Engineering Department, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. His research interests include evolutionary algorithms, fuzzy systems, neural networks, artificial immune systems and artificial muscles.

An Approach to Compose Viewpoints of Different Stakeholders in the Specification of Probabilistic Systems

Mahboubeh Samadi*
Faculty of Electrical and Computer Engineering, Shahid Beheshti University G. C. Tehran, Iran
mbh_samadi@yahoo.com
Hasan Haghighi
Faculty of Electrical and Computer Engineering, Shahid Beheshti University G. C. Tehran, Iran
h_haghighi@sbu.ac.ir

Abstract

Developing large and complex systems often involves many stakeholders each of which has her own expectations from the system; hence, it is difficult to write a single formal specification of the system considering all of stakeholders' requirements at once; instead, each stakeholder can specify the system from her own viewpoint first. Then, the resulting specifications can be composed to prepare the final specification. Much work has been done so far for the specification of non-probabilistic systems regarding viewpoints (or expectations) of different stakeholders; however, because of big trend to apply formal methods on probabilistic systems, in this paper, we present an approach to compose viewpoints of different stakeholders in the specification of probabilistic systems. According to this approach, different viewpoints are separately specified using the Z notation. Then, the resulting specifications are composed using some new operators proposed in this paper. We show the applicability of the presented approach by performing it on a known case study.

Keywords: Formal Methods, Formal Specification, Probabilistic Systems, Partial Models, Multiple Viewpoints.

1. Introduction

Developing large and complex systems often involves many stakeholders. However, each of these stakeholders has her own viewpoints (or expectations) when the system is specified. To collect requirements of all stakeholders, [1,2] proposed the viewpoint-oriented requirements engineering. From another point of view, modern systems are big and complex, resulting from assembling multiple components. "Components are designed by teams, working independently but with a common agreement on what the interfaces of each component should be [3]".

Considering both of the above mentioned cases, parallel and logical compositions should be done to produce a final specification of the system. "System specification through parallel composition is done by putting specifications of various components together. Logical composition (or merging), however, is used to merge viewpoints of different stakeholders to obtain the specification of a single component or system [4]".

As our review in section 2 shows, much of the related work has focused on parallel composition of partial specifications of non-probabilistic systems. Since there is a big trend to the formal specification and development of probabilistic systems, we present a formal method to compose viewpoints of different stakeholders when specifying a single probabilistic component or system (i.e., logical composition).

To achieve this goal, we first use the Z notation to specify viewpoints of different stakeholders separately.

As shown in the paper, Z schema calculus operations do not work on merging resulting specifications; hence, we define a new set of operators, including "m_conjunction", "m_disjunction" and "m_hiding" to compose specifications obtained after the first step. Names of operators begin with "m" which abbreviates for "merge".

Section 2 reviews the related work. In section 3 a probabilistic system is specified from viewpoints of different stakeholders. Section 4 first shows that Z schema calculus operations are not sufficient to compose specifications of different stakeholders and then presents a set of new operators. The applicability of our method is demonstrated using a known case study in section 5. And finally, section 6 is devoted to the conclusion and directions for future work.

2. Related Work

We categorize the works on the system specification from viewpoints of different stakeholders into two groups: specification of non-probabilistic systems and specification of probabilistic systems. There is little work on the latter category. Moreover, the presented methods are based on behavioural models.

Instead, much work has been done so far to specify non-probabilistic systems from viewpoints of different stakeholders. In [5], it is shown that partial models can be used to specify a system from viewpoints of different stakeholders. In this way, each model satisfies certain system requirements (from a certain stakeholder's point

of view). To make the final specification of the system, these models should be elaborated through both logical and parallel compositions in order to yield a system model that preserves the properties of the initial viewpoints altogether. In [8] and [9], a new operator, called conjunction, is introduced for the composition of different stakeholders' specifications. The behaviour of this operator is similar to the merge operator introduced in [5], but it is only defined for Modal Transition Systems (MTSs).

In [12] the modal interface framework, a unification of interface automata and modal specification is presented. The goal of this work is to compose specifications of interfaces from different viewpoints. The result of this work is a complete theory with a powerful composition algebra that includes operations such as conjunction (for requirements composition) and residuation (for components reuse that in addition assumes/guarantees contract-based reasoning [11]).

In [13], viewpoints are shown as partial specifications of functionality, written in Z but by different people, to be reconciled later. The focus of this work is on reconciliation and amalgamation of partial specifications and not the structure of these specifications themselves. By reconciliation, partial specifications become ready for the composition, and by amalgamation, real composition is done.

For collections of partial specifications to be meaningful, consistency between them has to be committed. In [15], it is described how to check consistency between partial specifications in Z, and how to ensure that different partial specifications of one system do not impose contradictory requirements; in [10] a solution to handle inconsistency between different specifications is introduced.

Besides the above mentioned work on non-probabilistic systems, a number of works have been done in the area of multi viewpoints specification of probabilistic systems. Interval Markov Chains (IMCs) and Constraint Markov Chains (CMCs) introduced in [16,17] can be used for non-functional analysis of multi viewpoints probabilistic systems [4]. In [4], it is shown that IMC is not a proper formalism for compositional specification. Thus, [4] and [17] introduce CMC for component based design of probabilistic systems. CMCs are a further extension of IMCs allowing rich constraints on the next-state probabilities from any state.

Larsen et. al. [18] further explore the influence of non-deterministic behaviour by mixing CMCs and MTSs and considering Probabilistic Automata (PA). In their model, state changes are additionally guarded by actions [4]. They present a specification theory for PAs, namely Abstract Probabilistic Automata (APA) which can serve as a specification theory for systems with both non-deterministic and stochastic behaviours. APA like any usable specification theory is equipped with a conjunction operator that allows combining multiple requirements into a single specification, and a composition operator that allows specifications to be combined structurally [17].

As described, there is not much work in the specification of probabilistic systems from viewpoints of different stakeholders. Moreover, the existing works on probabilistic systems use behavioural models to specify such systems while benefits of using well-known functional specification languages, such as Z, encourage us to specify probabilistic systems from viewpoints of different stakeholders using a Z-based formalism.

3. Specification of probabilistic systems from viewpoints of different stakeholders

In this section, we use the Z notation to write separate specifications (of a probabilistic system) describing viewpoints of different stakeholders. We propose our specification method through an illustrative example [17]. In this example, a customer and a manufacturer are considered as stakeholders of the system.

3.1 Example

Two parties, a customer and a manufacturer, are discussing a design of a relay for an optical telecommunication network. The relay should have several modes of operation, modelled by four dynamically changing properties and specified by atomic propositions a, b, c, and d as follows:

a: The Bit error rate is less than 1 per billion bits transmitted.

b: The Bit rate is higher than 10 Gbits/s.

c: Power consumption is less than 10 W.

d: The relay is not in the transmission mode (is in the standby state).

At first, informal specifications of the relay from customer's and manufacturer's viewpoints are presented.

- *Customer Specification:* In the initial state, the relay is in the standby mode (i.e., proposition "d" holds). Then, with a probability more than 0.7, it can move to state s_2 which is specified as $\{\{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$; this set means that in state s_2, at least two of properties "a", "b" and "c" hold, and "d" does not hold. With an unknown probability, the relay can move from the initial state to state s_3 specified as $\{\{a\}, \{b\}, \{c\}\}$; this means that in state s_3, exactly one of properties "a", "b" and "c" holds. The relay comes back to the initial state from states s_2 and s_3 with probability 1. Finally, there is no transition with the same source and destination (Figure 1).

- *Manufacturer Specification:* In the initial state, the relay is in the standby mode. Then with a probability more than 0.2, it can move to state s_3 where "a" and "d" do not hold. And with an unknown probability, it can move from the initial state to state s_2 where at least proposition "a" holds, and "d" does not hold. This relay comes back to

the initial state from states s_2 and s_3 with probability 1. Finally, there is no transition with the same source and destination (Figure 2).
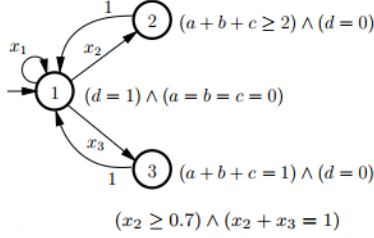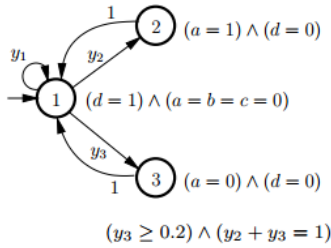

Fig. 1. The Customer's Specification

Fig. 2. The Manufacturer's Specification

3.2 Formal specification of the relay

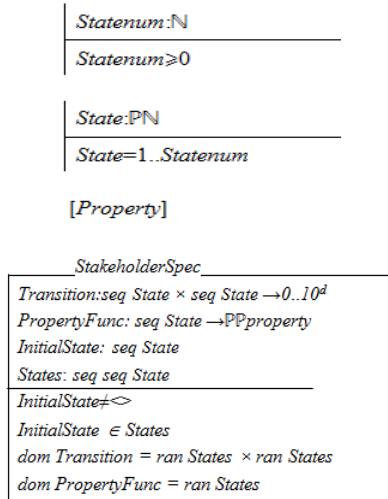Here is the formal specification of every probabilistic system from one stakeholder's point of view:

Fig. 3. The Stakeholder's Specification

Property shows a given type of properties (such as a, b, c, and d in our example) that could be true in each state. *StakeholderSpec*, as the state schema of the system, specifies all system states and their related properties and transitions from the stakeholder's point of view. It also shows the initial state of the system. Although each stakeholder prefers a set of desired bindings of the state schema, we assume that all of them agree on the initial state (*InitialState*) and the set of states (*States*). Since states in the final (composite) specification are combinations of states specified by each stakeholder, we consider a sequence of numbers (each number

corresponds to a state in one stakeholder's specification before combination) per each state, either it is composite or simple; for simple states, i.e., when we are considering the specification from a single stakeholder's viewpoint, this sequence has only one element. As an example of composite states, if the initial state specified by each of customer and manufacturer is <1>, the initial state in the final, composite specification will be shown as <1, 1>.

For two states s_i and s_j, *Transition* (s_i, s_j) is the probability of transition from s_i to s_j. Since floating-point numbers cannot be shown in the Z notation, transition probabilities are converted to natural numbers by multiplying them with 10^d. Thus, transition probabilities are shown as $0..10^d$. Considering all existing probabilities, d is the maximum number of digits to the right of the floating point. *PropertyFunc* is a function that assigns a set of sets of properties to each state; for example, consider set {{a, b}, {a, c}, {b, c}, {a, b, c}} for state s_2 in the customer's specification of the relay. The last two constraints in the schema guarantee that both *Transition* and *PropertyFunc* are defined on all of available states and anything else.

Regarding *StakeholderSpec* above, the formal specification of the relay from viewpoints of the customer and manufacturer is shown as *CustomerSpec* and *ManufacturerSpec* schemas, respectively.
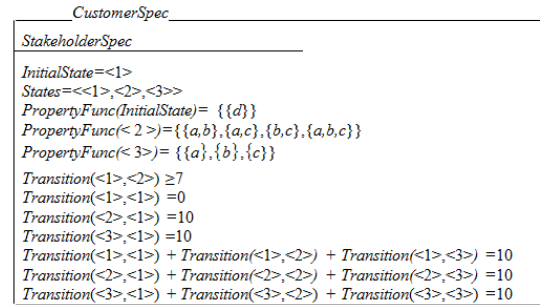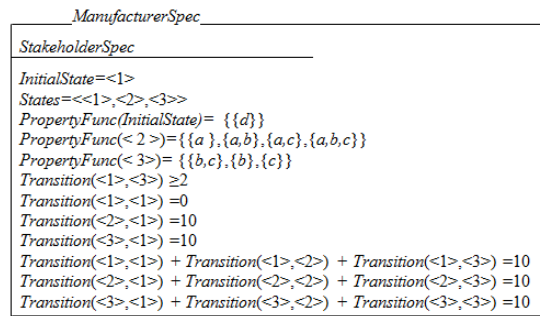

Fig. 4. The Customer's Specification

Fig. 5. The Manufacturer's Specification

4. Operators to compose viewpoints

As a simple example, suppose that we use the Z conjunction operator to compose *CustomerSpec* and *ManufacturerSpec*. The constraint part of the resulting schema will be *false* because, just as one example, *PropertyFunc(<2>)* is in the left hand of two equalities

with different right hand sides in *CustomerSpec* and *ManufacturerSpec*; the conjunction of these two equalities and thus the conjunction of schemas constraints will be *false*. Such a *false* constraint obviously satisfies neither the customer nor the manufacturer. Related to this example, it should be noticed that at least two of properties "a", "b" and "c" hold in state s_2 from the customer's viewpoint. On the other hand, at least proposition "a" holds in state s_2 from the manufacturer's viewpoint. Thus, we could have a composite state (in the final, composite specification) where "a" holds, and at least one of "b" and "c" holds. Such a state satisfies both the customer and the manufacture; similar examples can be given for the composition of other states and also transitions between states.

In summary, while we could have a composite specification satisfying both the customer and the manufacture, the Z conjunction operator is not able to generate such specification. Similarly, it can be shown that the disjunction operator in Z is not sufficient for merging specifications of different stakeholders in order to meet at least one of the viewpoints. Therefore, we are to define new operators to merge specifications of different stakeholders.

4.1   m_conjunction: meeting both viewpoints

The following definitions, that show how transition probabilities are determined in composite specifications, will be used when introducing the new operator m_conjunction.

Definition 1. For two states $< x_i >$ and $< x_j >$ in the first stakeholder's specification, if Transition($< x_i >, < x_j >) \geq a$, then for every states $< y_k >$ and $< y_{k'} >$ in the second stakeholder's specification for which Transition($< y_k >, < y_{k'} >) \geq 0$, we have $\sum_{y_k, y_{k'}}$ Transition($< x_i, y_k >, < x_j, y_{k'} >) \geq a$ in the composite specification.

Definition 2. For two states $< y_i >$ and $< y_j >$ in the second stakeholder's specification, if Transition($< y_i >, < y_j >) \geq a$, then for every states $< x_k >$ and $< x_{k'} >$ in the first stakeholder's specification for which Transition($< x_k >, < x_{k'} >) \geq 0$, we have $\sum_{x_k, x_{k'}}$ Transition($< x_k, y_i >, < x_{k'}, y_j >) \geq a$ in the composite specification.

Definition 3. For two states $< x_i >$ and $< x_j >$ in the first stakeholder's specification and for two states $< y_i >$ and $< y_j >$ in the second stakeholder's specification, if Transition $(< x_i >, < x_j >) = k$ and $Transition(< y_i >, < y_j >) = k'$, then Transition($< x_i, y_i >, < x_j, y_j >) = k * k'$.

The rationality behind Definition 1 is that, suppose based on the first stakeholder's viewpoint, the transition

probability from state $< x_i >$ to $< x_j >$ is greater than a. Now, based on the final, composite specification, the system should be able to move from states which start with $< x_i >$ to states which start with $< x_j >$ with a probability greater than a totally. In this way, the first stakeholder is satisfied by the final specification. A same reason can be given for Definition 2. Notice that two more definitions should be considered for $\leq$ the same as those presented for $\geq$.

Schema *CompositeSpec* below is used to compose specifications of different stakeholders in the form of *StakeholderSpec* (see subsection 3.2). We assume that we have two schemas of two stakeholders, called *FirstStakeholderSpec* and *SecondStakeholderSpec*. In addition, *StakeholderSpec* given in declaration part of *CompositeSpec* is the final, composite specification.

Since *FirstStakeholderSpec* and *SecondStakeholderSpec* have identifiers with the same name, in the declaration part of *CompositeSpec* their identifiers are renamed to avoid variable capturing: all identifiers are replaced with the same names but ended by F (for identifiers of *FirstStakeholderSpec*) and S (for identifiers of *SecondStakeholderSpec*). We use notation *FirststakeholderSpec[identifier/identifierF]* to show that one "F" is appended to the name of all identifiers of *FirstStakeholderSpec*. A similar notation is used for renaming identifiers of *SecondStakeholderSpec*.

Since the constraint part of *CompositeSpec* is almost long, we present it gradually via different parts. The informal description of each part is also given accordingly. For this reason, we do not draw the bottom line of *CompositeSchema* in Figure 6. As the first line of the constraint part, it is mentioned that the initial state of the system is obtained by concatenating the initial states of two stakeholders' viewpoints.
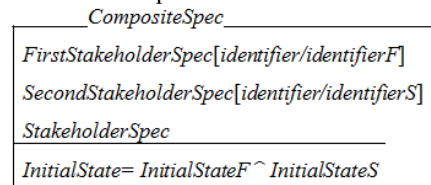

Fig. 6. Composite Specification – part 1

Besides the given equality for *InitialState*, *CompositeSpec* has the following constraints, too:

Fig. 7. Composite Specification – part 2

Before probing on the above predicate, we should mention that in the specification which supports viewpoints of both stakeholders, the set of states is as the Cartesian product of states in the specification of each stakeholder. More precisely, if sets of states in two stakeholder's specifications are $1..k_1$ and $1..k_2$, the set of states in the final specification is $(1..k_1) \times (1..k_2)$. Of course, the composition of two states will lead to an inconsistent state if the conjunction of their related properties is false, or in other words, the intersection of sets resulted from applying *PropertyFunc* to those two states is $\varnothing$. Inconsistent states will not appear in the final specification.

The predicate in Figure 7 says that probabilities of all transitions from\to an inconsistent state are 0. In other words, the system cannot move from\to inconsistent states; figures 8 and 9, in contrast, specify transition probabilities for consistent states. Based on the predicate given in Figure 8, if the transition probability from state s_1 to state s_2 in the specification of one stakeholder is in interval $[p_1,p_2]$, the total sum of transition probabilities from those states corresponding to s_1 to those states corresponding to s_2 in the final specification should be in interval $[p_1,p_2]$. This predicate is based on Definitions 1 and 2.

As another point, the utility function "sum" defined using an axiomatic definition in Figure 11 calculates the sum of elements in a sequence of natural numbers.

$\forall\ spindex_F: 1\ ..\ \#StatesF;\ index_F: 1\ ..\ \#StatesF;$

$\forall\ spindex_S: 1\ ..\ \#StatesS;\ p_1,p_2:\mathbb{N}\ |$

$ProperyFuncF(StatesF.\ spindex_F)\ \cap$

$PropertyFuncS(StatesS.\ spindex_S)\ \neq\varnothing\ \wedge$

$p_1 \leq TransitionF\ (StatesF.\ spindex_F, StatesF.\ index_F)\ \wedge$

$TransitionF\ (StatesF.\ spindex_F, StatesF.\ index_F) \leq p_2\ \bullet$

$\exists\ s:seq\mathbb{N}\ ;\ \forall\ i:1\ ..\ \#StateS\ \bullet$

$s.i = Transition(StatesF.\ spindex_F\ ^\frown StatesS.\ spindex_S,$

$StatesF.\ index_F\ ^\frown StatesS.i)\ \wedge\ p_1 \leq sum(s)\ \wedge\ sum(s) \leq p_2$

$\forall\ spindex_F: 1\ ..\ \#StatesF;$

$\forall spindex_S: 1\ ..\ \#StatesS;\ index_S: 1\ ..\ \#StatesS\ ;\ p_1,p_2:\mathbb{N}\ |$

$PropertyFuncS(StatesS.\ spindex_S)\ \cap$

$PropertyFuncF(StatesF.\ spindex_F)\ \neq\varnothing\ \wedge$

$p_1 \leq TransitionS\ (StatesS.\ spindex_S, StatesS.\ index_S)\ \wedge$

$TransitionS\ (StatesS.\ spindex_S, StatesS.\ index_S) \leq p_2\ \bullet$

$\exists\ s:seq\mathbb{N}\ ;\ \forall\ i:1\ ..\ \#StatesF\ \bullet$

$s.i = Transition(StatesF.\ spindex_F\ ^\frown StatesS.\ index_S,$

$StatesF.i\ ^\frown StatesS.\ index_S)\ \wedge\ p_1 \leq sum(s)\ \wedge\ sum(s) \leq p_2$

Fig. 8. Composite Specification – part 3

Unlike predicates in Figure 8 which consider transitions with variable probabilities, the following predicates regard transitions with constant probabilities (see Definition 3):

$\forall\ spindex_F:1\ ..\ \#StatesF;\ index_F: 1\ ..\ \#StatesF;$

$\forall spindex_S: 1\ ..\ \#StatesS;\ index_S: 1\ ..\ \#StatesS;\ p_1,p_2:\mathbb{N}\ |$

$PropertFuncF(StatesF.\ spindex_F)\ ^\frown$

$PropertyFuncS(StatesS.\ spindex_S) \neq\varnothing\ \wedge$

$TransitionF\ (StatesF.\ spindex_F, StatesF.\ index_F) = p_1\ \wedge$

$TransitionS\ (StatesS.\ spindex_S, StatesS.\ index_S) = p_2\ \bullet$

$Transition(StatesF.spindex_F\ ^\frown StatesS.\ spindex_S,$

$StatesF.\ index_F\ ^\frown StatesS.\ index_S) = p_1 \times p_2$

Fig. 9. Composite Specification – part 4

In the following predicate, it is emphasized that the sum of probabilities of transitions from one state in the final specification should be 1 (or in fact 10^d since we multiplied all probabilities with 10^d); the utility function "bind" defined using an axiomatic definition in Figure 11 constructs a sequence through the concatenation of all sequences existing in a sequence of sequences.

$\exists\ s:seqseq\mathbb{N}\ ;\ \forall\ spindex_F:1\ ..\ \#StatesF;\ i: 1\ ..\ \#StatesF;$

$\forall spindex_S: 1\ ..\ \#StatesS; j: 1\ ..\ \#StatesS\ \bullet$

$(s.i).j = Transition(StatesF.\ spindex_F\ ^\frown StatesS.\ spindex_S,$

$StatesF.i\ ^\frown StatesS.j)\ \wedge\ sum(bind(s)) = 10^d$

Fig. 10. Composite Specification – part 5

$$sum: seq\mathbb{N} \rightarrow \mathbb{N}$$
$$sum<>=0$$
$$\forall\ s:seq\mathbb{N}\ |\ s \neq <>\ \bullet\ sum(s) = head(s) + sum(tail(s))$$

$$bind: seqseq\mathbb{N} \rightarrow seq\mathbb{N}$$
$$bind<>=<>$$
$$\forall s:seqseq\mathbb{N}\ |\ s \neq <>\ \bullet\ bind(s) = head(s)\ ^\frown bind(tail(s))$$

Fig. 11. Utility Functions

The following predicate says that state properties in the final specification are the conjunction of state properties specified by each stakeholder.

$\forall\ si: 1..\#StatesF; sj: 1\ ..\ \#StatesS \bullet PropertyFunc(StatesF.si\ ^\frown StatesS.sj) =$

$PropertyFuncF(StatesF.\ si) \cap PropertyFuncS(StatesS.sj)$

Fig. 12. Composite Specification – the last part

Now, since the final specification should only consist of composed states and related transitions, specifications of the first and the second stakeholders should be hidden. Consequently, the new operator for merging viewpoints of two stakeholders (in order to satisfy both of them) is specified as follows:

$FirstStakeholderSpec\ \wedge_{VoS} SecondStakeholderSpec ==$

$\exists\ FirstStakeholderSpec[identifier/identifierF],$

$SecondStakeholderSpec[identifier/identifierS]\ \bullet\ CompositeSpec$

Fig. 13. m_conjunction of Two Schemas

Here, $\wedge_{VoS}$ is the symbol of m_conjunction in which VoS abbreviates for "Viewpoints of Stakeholders".

4.2 m_disjunction: meeting at least one of viewpoints

Sometimes, it is important that the final specification satisfies the concerns of at least one stakeholder. We introduce operator "m_disjunction" to support this requirement. In the final specification, the states are all of the states specified by all stakeholders. Since states in two specifications may include identical numbers, and we are going to consider all states in the final specification, schema *Rename* in Figure 14 specifies the change of numbers used in states of the second specification. This change should be done before we merge the two specifications.

$$
\begin{array}{l}
\underline{\quad Rename \quad} \\
FirstStakeholderSpec[identifier/identifierF] \\
\Delta SeconStakeholderSpec[identifier/identifierS] \\
\rule{4cm}{0.4pt} \\
StatesS' = Add2ToAllElems\ (StatesS) \\
TransitionS' = Add2ToAllTStates\ (TransitionS) \\
PropertyFuncS' = Add2ToAllPStates\ (PropertyFuncS)
\end{array}
$$

$$FinalCompositeSpec = Rename\,\mathring{\,}\,LeastCompositeSpec$$

Fig. 14. Change of Numbers in the Second Specification

Three functions *Add2ToAllElems, Add2ToAllTStates* and *Add2ToAllPStates* are supposed to add number "2" at the beginning of numbers used in the states of the second specification. For example, state <1> is changed to <21>. Due to the space limitation, we do not define these functions here. The final specification is obtained by sequential composition of *Rename* and *LeastCompositeSpec* specified below.

Schema *LeastCompositeSpec* specifies the composition of stakeholder specifications using m_disjunction. This schema includes the schemas of two stakeholders. Similar to what we did for m_conjunction, we rename identifiers of *FirstStakeholderSpec* and *SecondStakeholderSpec* to avoid variable capturing here.

$$
\begin{array}{l}
\underline{\quad LeastCompositeSpec \quad} \\
FirstStakeholderSpec[identifier/identifierF] \\
SecondStakeholderSpec[identifier/identifierS] \\
StakeholderSpec \\
\rule{4cm}{0.4pt} \\
InitialState = \ <0> \\
States = <InitialState> \cap StatesF \cap StatesS \\
Transition = TransitionF \cup TransitionS \cup \\
\{((InitialState,\ InitialStateF),\ 0.5 \times 10^d)\} \cup \\
\{((InitialState,\ InitialStateS),\ 0.5 \times 10^d)\} \\
PropertyFunc = PropertyFuncF \cup PropertyFuncS \cup \{(InitialState,\ \varnothing)\}
\end{array}
$$

Fig. 15. *LeastCompositeSpec* Schema

We define new state <0> as the initial state. This state has no property (see the last line of the constraint part). Since the states in the composite system should be all of the states specified by both stakeholders, states are specified as the concatenation of states specified by the first and the second stakeholder and also the new defined state (i.e., <0>). To consider the two viewpoints in the

same way, we add two new transitions with probability 0.5 from the new initial state: one to the initial state in *FirstStakeholderSpec* and the other one to the initial state in *SecondStakeholderSpec*.

Finally, the new operator for composing viewpoints of two stakeholders (in order to meet at least one of them) is specified as follows ($\vee_{\text{VoS}}$ is the symbol of m_disjunction):

$$
\begin{array}{l}
FirstStakeholderSpec \vee_{VoS} SecondStakeholderSpec == \\
\exists FirstStakeholderSpec[identifier/identifierF] \\
, SecondStakeholderSpec[identifier/identifierS] \\
\bullet FinalCompositeSpec
\end{array}
$$

Fig. 16. m_disjunction of Two Schemas

4.3 m_hiding: hiding states

Sometimes it is required to hide one state before using the system specification from one stakeholder's viewpoint. Here are some examples:

- A stakeholder would not rather see one state of the system that another stakeholder specifies.
- To apply some change to a given specification in order to make it reusable in another situation.
- Each stakeholder may change her specification according to her new viewpoint only by hiding states.
- Before composing specifications using the m_conjunction operator, it may be necessary to change one or both specifications by hiding states.

Hide schema is as follows:

$$
\begin{array}{l}
\underline{\quad Hide \quad} \\
\Delta StakeholderSpec \\
HiddenState : StateSet \\
\rule{4cm}{0.4pt} \\
HiddenState \neq InitialState \\
\exists x: 1..\#States.Transition(HiddenState, States.x) > 0 \\
HiddenState \neq <> \\
\forall x: 1..\#States\ |States.x \neq HiddenState\ .\ States.x \in ran\ States' \\
\#States' = \#States - 1
\end{array}
$$

Fig. 17. *Hide* Schema - part 1

In the declaration part of this schema, the state being hidden is specified as *HiddenState*. In the constraint part, it is mentioned that the hidden state should not be the initial state of the specification. In addition, it should be the source of at least one transition. These two constraints are given to guarantee that no probability value is missed after removing the hidden state. The last two predicates describe removing the hidden state.

$$\forall x:1..\#States\ .\ Transition'(States.x, HiddenState) = 0$$

$$\forall x:1..\#States\ .\ Transition'(HiddenState, States.x) = 0$$

Fig. 18. *Hide* Schema – part 2

The above predicate says that every transition whose source or destination is the hidden state should be discarded. Instead, for arbitrary states x and y, if there is one transition from x to the hidden state and one

transition from the hidden state to y, the multiplication of probabilities of these two transitions should be added to the current probability of the transition from x to y. This constraint is described in Figure 19.

$$\forall x,y: 1 .. \#States;\ p_1,p'_1,p_2,p'_2: \mathbb{N} \bullet$$

$$p_1 \leq Transition(States.\ x, HiddenState) \land$$

$$Transition(States.\ x, HiddenState) \leq p_2 \land$$

$$p'_1 \leq Transition(HiddenState, States.y) \land$$

$$Transition(HiddenState, States.y) \leq p'_2 \longrightarrow$$

$$p_1 \times p'_1 + Transition(States.x,\ States.y) \leq Transition'(\ States.\ x, States.y)$$

$$\land\ Transition'(\ States.\ x, States.y) \leq p_2 \times p'_2 + Transition\ (States.x, States.y)$$

Fig. 19. *Hide* Schema – the last part

At last, the new operator for hiding a state from one specification is specified as follows:

$$StakeholderSpec \setminus_{VoS} HiddenState = \ = \exists\ Stakeholder \bullet Hide \setminus HiddenState$$

Fig. 20. m_hiding Definition

Here, $\setminus_{VoS}$ is the symbol of m_hiding.

5. Case Study

In subsection 3.2, a relay for an optical telecommunication network was specified from viewpoints of different stakeholders, i.e., a customer and a manufacturer. The constraint part of *CustomerSpec* $\land_{VoS}$ *ManufacturerSpec* below specifies states properties and transitions probabilities in the relay from both stakeholders' viewpoints. Properties of each state are conjunction of properties specified by each stakeholder. It is worth noting that inconsistent states and their relevant transitions have not been shown in the schema.

```
___ CustomerSpec ∧_VoS ManufacturerSpec _____

  StakeholderSpec
 _____

  InitialState=<1,1>
  States=<<1,1>,<2,2>,<3,3>,<2,3>,<3,2>,<1,2>,<1,3>>
  PropertyFunc(<1,1>)={{d}}
  PropertyFunc(<2,2>)={{a,c},{a,c},{a,b,c}}
  PropertyFunc(<2,3>)={{b,c}}
  PropertyFunc(<3,2>)={{a}}
  PropertyFunc(<3,3>)={{c},{b}}
  Transition(<1,1>,<2,2>)+ Transition(<1,1>,<2,3>)>7
  Transition(<1,1>,<2,3>)+ Transition(<1,1>,<3,3>)>2
  Transition(<1,1>,<1,1>)+ Transition(<1,1>,<1,2>)+ Transition(<1,1>,<1,3>)=0
  Transition(<1,1>,<1,1>)+ Transition(<1,1>,<2,1>)+ Transition(<1,1>,<3,1>)=0
  Transition(<2,2>,<1,1>)=10
  Transition(<2,3>,<1,1>)=10
  Transition(<3,2>,<1,1>)=10
  Transition(<3,3>,<1,1>)=10
  Transition(<1,1>,<2,2>)+ Transition(<1,1>,<2,3>)+
  Transition(<1,1>,<3,3>)+ Transition(<1,1>,<3,2>)=10
```

Fig. 21. *CustomerSpec* $\land_{VoS}$ *ManufacturerSpec*

Sometimes, it is required that the concerns of at least one stakeholder are satisfied. To achieve this goal, m_disjunction of schemas is useful. Figure 22 indicates the application of m_disjunction to *CustomerSpec* and *ManufacturerSpec*.

```
___ CustomerSpec ∨_VoS ManufacturerSpec _____

  StakeholderSpec
 _____

  InitialState=<0>
  States=<<0>,<1>,<2>,<3>,<21>,<22>,<23>>
  PropertyFunc<0>={}
  PropertyFunc<1>={{d}}
  PropertyFunc<2>={{a,b},{a,c},{b,c},{a,b,c}}
  PropertyFunc<3>={{a},{b},{c}}
  PropertyFunc<21>={{d}}
  PropertyFunc<22>={{a},{a,b},{a,c}{a,b,c}}
  PropertyFunc<23>={{b,c},{b},{c}}
  Transion(<0>,<1>)=5
  Transion(<0>,<21>)=5
  Transition(<1>,<2>) ≥7
  Transion(<1>,<1>)=0
  Transion(<2>,<1>)=10
  Transion(<3>,<1>)=10
  Transition(<1>,<1>)+Transition(<1>,<2>) +Transition(<1>,<3>)=10
  Transition(<2>,<1>)+Transition(<2>,<2>) +Transition(<2>,<3>)=10
  Transition(<3>,<1>)+Transition(<3>,<2>) +Transition(<3>,<3>)=10
  Transition(<21>,<23>)≥2
  Transition(<21>,<21>)=0
  Transition(<22>,<21>)=10
  Transition(<21>,<21>)+Transition(<21>,<22>) +Transition(<21>,<23>)=10
  Transition(<22>,<21>)+Transition(<22>,<22>) +Transition(<22>,<23>)=10
  Transition(<23>,<21>)+Transition(<23>,<22>) +Transition(<23>,<23>)=10
```

Fig. 22. *CustomerSpec* $\lor_{VoS}$ *ManufacturerSpec*

Besides the new initial state, i.e., <0>, the final specification consists of all states in the initial specifications. Also, regardless of the new transitions starting from the new initial state, transition probabilities remain unchanged. The resulting specified relay can behave like customer's specification or manufacturer's specification.

Sometimes, it is required to hide one state before using the system specification from one stakeholder's viewpoint. Figure 23 indicates the application of m_hiding to *CustomerSpec* to hide state <2>.

```
___ CustomerSpecAfterHiding _____

  Stakeholderspec
 _____

  InitialState=<1>
  PropertyFunc<1>={{d}}
  PropertyFunc<3>={{a},{b},{c}}
  Transition(<1>,<1>) ≥7
  Transition(<3>,<1>) =10
  Transition(<1>,<1>) + Transition(<1>,<3>) =10
  Transition(<3>,<1>) + Transition(<3>,<3>) =10
```

Fig. 23. Hiding State <2> in *CustomerSpec*

6. Conclusions

In this paper, an approach to compose viewpoints of different stakeholders in the specification of probabilistic systems was presented. The main contribution of this approach is introducing three new schema operators to manipulate specifications written by different stakeholders. In the extended version of this paper, we are going to formally prove that the proposed operators are sound. Also, as another future work, we will present an approach to specify component based probabilistic systems. To achieve this goal, specifications of different stakeholders should be first merged to construct the specification of each component (logical composition), and then specifications of different components should be combined to construct the specification of the whole system (parallel composition).

References

[1] B. Nuseibeh, J. Kramer, A. Finkelstein, "A Framework for Expressing the Relationships between Multiple Views in Requirements Specification," *In IEEE Transaction on Software Engineering*, 1994, vol. 20, no. 1, pp: 760-773.

[2] G. Kotonya, and I. Sommerville, "Requirements Engineering with Viewpoints," *Journal in Software Engineering*, 1996, vol. 11, no. 2, pp: 58-66.

[3] S. Rui-feng, Y. Chao, X. Jie, "Acquire Multi-Viewpoint from Domain," In *3rd International Conf. on Advance Comp. Theory and engineering*, 2010, vol. 4, pp.598-602.

[4] B. Caillaud, B. Delahaye, K.G. Larsen, A. Legay, M. Pederson, A. Wasowski, "Compositional Design Methodology with Constraint Markov Chains," *In 7th International Conference on the Quantitative Evaluation of Systems*, 2010, pp 123-132.

[5] A. Benveniste, B. Caillaud, R. Passerone, "Multi-Viewpoint State Machines for Rich Component Models," In *Model-Based Design of Heterogeneous Embedded Systems*, Pieter Mosterman, Gabriela Nicolescu (eds.), 2009.

[6] S. Uchitel, and M. Chechik, "Merging Partial Behavioural Models," In *12th ACM International symp. on foundation of soft. Engineering*, 2004, vol. 29, no. 6, pp. 43-52.

[7] G. Brunet, M. Chechik, S. Uchitel, "Properties of Behavioural Model Merging," In *14th International Conference on Formal Methods*, 2006, pp. 98-114.

[8] K.G. Larsen, B. Steffen, C. Weise, "A Constraint Oriented Proof Methodology Based on Modal Transition Systems," In *first International workshop on Tools and Algorithms for the Construction and Analysis of Systems*, 1995, pp. 17-40.

[9] M. Huth, R. Jagadeesan, and D. Schmidt, "A Domain Equation for Refinement of Partial Systems," In *Mathematical Structures in Computer Science*, 2004, vol. 4, no. 3.

[10] C, Anthony, D. Gabbay, A. Hunter, J. Kramer, and B. Nuseibeh, "Inconsistency Handling in Multiperspective Specifications," In *IEEE Transaction on Software Engineering*, 1994, vol 20.no X A1 Gl ST.

[11] D. Fischbein, G. Brunet, N. D'Ippolito, M. Chechik, and S. Uchitel, "Weak Alphabet Merging of Partial Behaviour Models," In *ACM Transactions on Software Engineering and Methodology*, 2011, vol. 21, pp: 1-49.

[12] A. Benveniste1, B. Caillaud1, and A. Ferrari, Multiple Viewpoint Contract-Based Specification and Design. In *FMCO*, Springer, 2008, , pp: 200–225.

[13] JB. Raclet, A. Benveniste, A. Legay, and et al, "A Modal Interface Theory for Component-based Design," In *Fundamenta Informaticae*, 2010.

[14] D. Jackson, "Structuring Z Specifications with Views," In *ACM Transactions on Software Engineering and Methodology*, 1995, vol. 4.

[15] M. Ainsworth, AH. Cruickshank, L.J. Groves and P.J.L. Wallis, "Formal Specification via Viewpoints," In *Proc. 13th New Zealand Computer Conference*, 1993.

[16] E. Boiten, J. Derrick, H. Bowman, M. Steen, "Constructive Consistency Checking for Partial Specification in Z," In *Science of Comp. Prog.*, 1995, vol. 35, pp: 29-75.

[17] B. Delahaye, K.G. Larsen, A. Legay, M. Pedersen, A. Wasowski, "Decision Problems for Interval Markov chains," In *5th International Conference on Language and Automata Theory and Applications*, 2011, pp. 274-285.

[18] B. Caillaud, B. Delahaye, K.G. Larsen, A. Legay, M. Pederson, A. Wasowski, "Constraint Markov Chains," *Journal In Theoretical Computer Science*, 2011, vol. 412, no. 5, pp.4373-4404.

[19] B. llaud, B. Delahaye, K.G. Larsen, A. Legay, M. Pederson, A. Wasowski, "New Results on Abstract Probabilistic Automata," In *11th International Conference on Application of concurrency to system design*, 2011, pp. 118-127.

Mahboubeh Samadi received her B.Sc. and M.Sc. degrees in software engineering from the Faculty of Electrical and Computer Engineering, Shahid Beheshti University, Tehran, Iran, in 2010 and 2012, respectively. Her research interests include software engineering and formal methods.

Hasan Haghighi received his B.Sc., M.Sc., and Ph.D. degrees from the Computer Engineering Department, Sharif University of Technology, Tehran, Iran, in 2002, 2004, and 2009, respectively. Since 2009, he has been with the Faculty of Electrical and Computer Engineering, Shahid Beheshti University, Tehran, Iran. His research interests include formal methods, software engineering, software architecture and software test.

An Improved Method for TOA Estimation in TH-UWB System considering Multipath Effects and Interference

Mahdieh Ghasemlou*
Department of Telecommunications Engineering, Islamic Azad University Tehran South branch, Tehran, Iran
mahdieh_ghasemlou@yahoo.com
Saeid Nader-Esfahani
School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran
nader@ut.ac.ir
Vahid Tabataba-Vakili
School of Electrical Engineering, Iran University of Science & Technology, Tehran, Iran
vakily@iust.ac.ir

Abstract

UWB ranging is usually based on the time-of-arrival (TOA) estimation of the first path. There are two major challenges in TOA estimation. One challenge is to deal with multipath channel, especially in indoor environments. The other challenge is the existence of interference from other sources. In this paper, we propose a new method of TOA estimation, which is very robust against the interference. In this method, during the phase of TOA estimation, the transmitter sends its pulses in random positions within the frame. This makes the position of the interference relative to the main pulse to be random. Consequently, the energy of interference would be distributed, almost uniformly, along the frame. In energy detection methodes, a constant interference along the frame does not affect the detection of arrival time and only needs the adjustment of the threshold. Simulation results in IEEE.802.15.4a channels show that, even in presence of very strong interference, TOA estimation error of less than 3 nanoseconds is feasible with the proposed method.

Keywords: Threshold, Interference, TOA, Ranging.

1. Introduction

Owing to its high time resolution, impulse radio-ultra wideband (IR-UWB) technology is an excellent signaling for accurate wireless localization used in many security, medical, search and rescue, and military applications [1]. Ranging with the use of UWB systems is usually done from the time-of-arrival (TOA) of the first path of the received signal; in these systems, the distance between the receiver and transmitter is measured based on the estimation of the signal propagation delay, which can be challenging in the presence of multipath and interference [2-3].

When channel parameters are unknown like in the case of multipath environment, the estimation of the TOA is dependent upon the channel estimation [4-5]; in this case, channel coefficients and delays are estimated using, for example, the maximum likelihood method. However, the very high complicacy of this method due to the high number of paths in a real multi-path channel makes its implementation limited. Coherent methods like matched filtering [6] are also costly to be implemented, due to their need for high rate sampling. To reduce complexity of UWB systems, non-coherent methods such as energy detector owing to their sub-Nyquist sampling, and consequently their low implication cost, are in spotlight of attention [7-8]. Algorithms based on energy detector are generally divided into two categories including

i) maximum energy selection (MES), and ii) threshold crossing (TC) algorithms.

MES-based TOA algorithms depend on the selection of the maximum amount of energy, and hence, have limited application in real multipath channels in which the first path is not usually the strongest one [2]. In TC-based TOA algorithms, receiver samples are compared with a threshold, and the TOA is estimated as the first sample crossing the threshold. The main feature of threshold-based algorithms is their analog implementation, which has attracted substantial attention in some applications like wireless sensor networks needing low-consumption devices [6].

The choice of the threshold strongly influences the performance of the TOA estimation. There are different methods for setting the threshold; it can be selected on the basis of noise level, regardless of channel properties and the energy of the received signal [9-10]; the threshold can also be defined on the basis of a constant normalized value between the lowest and highest amounts of energy samples [8]. Better approaches have been proposed for determining the normalized threshold on the basis of the received signal statistics such as its kurtosis [11], and the comparison between kurtosis and skewness results in the fact that skewness is more appropriate for accurate TOA estimation [12], and in [13] also the four statistical parameters including kurtosis skewness, maximum slope and standard deviation are compared. Since the maximum

* Corresponding Author

slope and skewness are more sensitive to SNR changes, the common criterion for determining the thresholds has been calculated using skewness and maximum slope.

In [14], a new method has been proposed to improve the estimation accuracy based on the received signal characteristics, that is based on determine 2 thresholds. Threshold in the first step is calculated based on the Gaussian distribution of noise samples, and the second one is based on the times energy samples in different threshold frames are more than the threshold in the first step. In [15], the first step threshold is calculated based on the chi -square distribution. Although it is expected that the chi -square distribution is more convenient for the noise samples, but the results show that since the threshold in the first step is not appropriate, so there is not much difference between the Gaussian and chi square distributions. So, as it was mentioned in [14] the first step requires a more appropriate threshold.

A new method has been proposed based on a rank test, which suffers from high complexity, especially when the number of frames increases [16]. Another thresholding method has also been introduced on the basis of the channel impulse response and real environment measurements [17].

Ranging through non-coherent receivers in the presence of interference is difficult, as the system may estimate the moment at which the interference is received as the real TOA. Many methods have been so far proposed for reducing interference; for example, nonlinear filters have been suggested to be used for lowering interference effects [18-19]; however, these method result in substantial errors in the presence of strong interferences.

In a recent conference paper [20], we have put forward a new method for estimating the TOA in a multipath channel and in the presence of interference. In this method, the transmitter sends its pulse not in a constant but a random position within the frame; therefore, the position of the interference relative to the main pulse is always changing, making the interference energy evenly distributed in different distances from the main pulse; such an interference will not have any effect in detecting the moment the pulse is entered, TOA, and it is just needed to increase the threshold to the level of this interference. To overcome the multipath problem of the channel, we have made use of serial backward search for multiple cluster (SBSMC) algorithm [19].

In [20] we introduced an approximate formula for the threshold, using a simple model. The present work is an extension of [20], in that a more accurate threshold is introduced by using an exponential profile for the channel model, according to IEEE 802.15.4a. This threshold can be adjusted for different environments by using the cluster decay factors of that environment.

The rest of the paper has been organized as follows; in section 2, the model of the UWB system with time hopping (TH) and the structure of the receiver have been discussed; section 3 has been devoted to the ED-based

algorithms for the TOA estimation as well as our proposed method for threshold determination. The performance of the present method has been compared with other methods in section 4, followed by a conclusion in section 5.
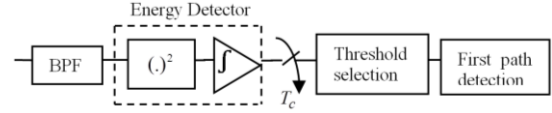

Fig. 1 Threshold-based TOA estimator with ED scheme

2. System Model and Receiver Structure

2.1 The Shape of the Signal Wave and the Structure of the Receiver

IEEE 802.15.4a [21] is an international standard modeling UWB channels with the use of Saleh-Valenzuela model. In TH-IR signaling, each symbol within a period of T_{sym} is divided into N_f number of time intervals of T_f the so-called frames. Each frame is also divided into N_c intervals of T_c called chips. The TH codes are randomly denoted by $c_j \in \{1, 2, …, N_h\}$ where N_h represents maximum hopping within a frame, and generally, $N_h \leq N_c$; here, we have considered $N_h = N_c$.

Fig. 1 shows a design of the ED-based TOA estimator [11]. To remove the out-of-band noise, the received signal first passes through a band-pass filter (BPF) with the bandwidth B and center frequency f_0. If we consider the first user as the desired user, the output of the BPF in the presence of interference can be written as [22]:

$$r(t) = s(t) + i(t) + n(t) \qquad (1)$$

$$s(t) = \sum_{n=1}^{N_{sym}} \sum_{k=1}^{N_f} w^{(1)}(t - c_k^{(1)}T_c - (k-1)T_f - (n-1)T_{sym}) \qquad (2)$$

$$s(t) = \sum_{n=1}^{N_{sym}} \sum_{k=1}^{N_f} w^{(1)}(t - c_k^{(1)}T_c - (k-1)T_f - (n-1)T_{sym}) \qquad (2)$$

and

$$w^{(u)}(t) = \sqrt{\frac{E_s^{(u)}}{N_f}} \sum_{l=1}^{L} \alpha_l^{(u)} P\left(t - \tau_l^{(u)}\right) \qquad (3)$$

where L is the number of the multipath components for the uth user; $\alpha_l^{(u)}$ and $\tau_l^{(u)}$ are respectively amplitudes and delays of the lth multipath component of the uth user, and $\tau_1^{(1)}$ is estimated as the TOA. Without loss of generality, the normalization $\sum_{l=1}^{L} E\{[\alpha_l^{(u)}]^2\} = 1$ can be considered for all users. N_{sym} is the number of the symbols, and $E_s^{(u)}$ represent average received energy per symbol. Component $i(t)$ is the interference, and $n(t)$ is the additive white Gaussian noise (AWGN) with zero mean and the two-sided power spectral density $\frac{N_0}{2}$. No modulation is considered for the ranging process.

When U users exist, so that each user transmits N_{sym} symbols at each time, i(t) can be expressed as:

$$i(t) = \sum_{u=2}^{U} \sum_{n=1}^{N_{sym}} \sum_{k=1}^{N_f} w^{(u)}(t - c_k^{(u)} T_c$$
$$- (k-1)T_f - (n-1)T_{sym}) \qquad (4)$$

After passing the filter, the received signal is entered the ED which is usually composed of a non-linear squaring element and an integrator. A sample is taken from the output of the integrator along the received signal after each T_c seconds. Therefore, energy samples of z[k] are obtained as follows:

$$z[k] = \int_{(k-1)T_c}^{kT_c} |r(t)|^2 dt \qquad \text{·}k = 1,...,N_b \qquad (5)$$

where $N_b = N_{sym}.n_b$ and $n_b = N_f.N_c$.

2.2 Energy Matrix of TH-IR

In TH-IR, the energy samples of z[k] are regulated according to the transmitted TH code, making the energy matrix of the dimensions of $(N_f.N_{sym} \times N_c)$ as follows [23]:

$$z[k(j),n] = z[n + (k-1)N_f \cdot N_c + (j-1)N_c + c_j^{(1)}] \qquad (6)$$

Where $k(j) = N_f.(k-1) + j$, $j \in \{1, 2, ..., N_f\}$, $1 \le k \le N_{sym}$ and $1 \le n \le N_c$.

2.3 Nonlinear filtering

To reduce the interference, many methods have been proposed to be applied to the energy matrix [19,21]. Therefore, before the TOA estimator, a filtering can be applied to the samples. The most common approach is to use an averaging filter obtaining the average of the samples in each column; therefore,

$$z[j,n] = \frac{1}{N_{sym}} \sum_{k=1}^{N_{sym}} z[k(j),n] \qquad (7)$$

We particularly use a median filter, and thus have [23]:

$$z[j,n] = median\{z[j,n], z[j+1,n],..., z[j+w-1,n]\} \qquad (8)$$

Where $\in \{1, 2, ..., N_f\}$, $1 \le n \le N_c$ and w is the length of the filter.

Since the nonlinear filtering performance is strongly reduced in the presence of the interference [19], in our proposed method, the averaging filter is first applied to each column of the matrix, and then, the error in the TOA estimation is removed through choosing a new threshold. At the next stage, the energy samples of the transmitted pulses within different frames are summed together. Therefore:

$$z[n] = \sum_{j=1}^{N_f} z[j,n] \qquad (9)$$

3. ED-based TOA Estimation Algorithms

There are many ED-based algorithms for the TOA estimation used for detecting the first path; the simplest algorithm is the MES choosing the first maximum energy:

$$\hat{t}_{MES} = [\arg\max\{ z[n]\} + 0.5]T_c = (n_{max} + 0.5)T_c \qquad (10)$$
$$1 \le n \le N_c$$

where the TOA is considered as the center of the integral interval. It is possible that the maximum value of the energy is not the first, giving rise to errors even in the high signal-to-noise ratio (SNR) [8]. Therefore, the estimation of the TOA is done on the basis of the TC, and the received values are compared with a suitable threshold ξ; in this case, the TOA is estimated as follows:

$$\hat{t}_{TC} = [\min\{n|z[n] > \xi\} + 0.5]T_c \qquad (11)$$

Table 1: Comparison of the cluster decay factors for different environments [21]

Environment	Residential		Office		Outdoor	
	LOS	NLOS	LOS	NLOS	LOS	NLOS
Channel Model	CM1	CM2	CM3	CM4	CM5	CM6
Γ (ns)	22.61	26.27	14.6	19.8	31.7	104.7

3.1 The Proposed Method

In the common method, the transmission position of the main pulse is fixed within the frame. If the interference along different symbols is also occurred at a constant position of the frame, the TOA will be strongly affected after averaging over different frames; and if the interference pulse starts before the main pulse, it will be estimated as the main entrance time, leading to large errors. The main idea of the present work is to get a constant interference energy along the frame after averaging; a constant interference along the frame does not cause any error in detecting the starting time of the pulse, and it is sufficient to take the effect of this constant interference just in the chosen threshold; to this end, it is needed to consider the TH code of the users along different symbols to be pseudo-random, and to use $N_h = N_c$; then, owing to the point that the interference position with respect to the main pulse is always changing in different symbols, the interference energy after averaging over the symbols will be evenly distributed at different distances from the main pulse, and thus, will be approximately constant. The assumption of $N_h = N_c$ results in interference between the frames, which should be taken into account in determining the threshold.

3.2 The Proposed Threshold Determination

Determining an appropriate threshold is of vital significance, and strongly influences the estimation of the TOA. None of the threshold determination methods proposed so far [6,8,9,11] can provide us with an appropriate threshold in the presence of the interference, as the system considers the arrival time of the interference received before the main pulse as the TOA. However, in

our proposed method, if the number of the symbols N_{sym} is large enough, the sum of the interference and noise along the frame will be approximately constant, and thus, can be considered as the threshold. Our simulations indicate that about 1000 symbols suffice to have the summation of the interference and noise energies constant; this constant value is indeed the expectation value of the sum of the energies of the noise, and the inter frame interference (IFI) and multi user interferences (MUIs):

$$\xi = \mu_N + \mu_{IFI} + \mu_{MUI} \qquad (12)$$

where μ_N , μ_{IFI} , and μ_{MUI} are respectively the expectation values of the energies of the noise, IFI, and MUI within each chip. It can be proven that [24]:

$$\mu_N = N_f \cdot B \cdot N_0 \cdot T_c \qquad (13)$$

$$\mu_{IFI} = (3 + \frac{1}{e^{T_c/\Gamma} - 1}) \frac{E_s}{N_c^2} \qquad (14)$$

$$\mu_{MUI} = \sum_{u=2}^{U} \frac{E_s^{(u)}}{N_c} \qquad (15)$$

where Γ is the exponential decay constant of the clusters in the IEEE 802.15.4a model. This coefficient has been tabulated in Table 1 for different environments [21]. N_0 is the noise density. The coefficient N_f in μ_N is because of the summation done over N_f energy samples of the noise within the frames of a symbol. $\sum_{u=2}^{U} E_s^{(u)}$ is the sum of the energies of the interfering users in each symbol, and $E_s^{(1)}$ is the energy of the desired user, which for the sake of simplicity is indicated as $E_s^{(1)} = E_s$.

3.3 SBSMC Algorithm

The multipath components received in UWB channels enter the receiver in many clusters, and the cluster of the first path may not be the strongest. To detect the arrival time of the first cluster, we have made use of the SBSMC algorithm [19] which is on the basis of detecting the largest sample and backward element-by-element search within a window of the length of w_{sb}; in this algorithm, the number of K samples, which are the noise and interference samples, before the detected sample should be smaller than the threshold. w_{sb} is calculated on the basis of the time delay at which the largest path is received after the arrival of the first path, which is typically 60 ns [18]. Hence, the TOA is estimated as follows [19]:

$$\hat{t} = [\max \arg \in \{n_{max}, ..., n_{max} - w_{sb}\} | z[n] > \xi \text{ and}$$
$$\max\{z[n-1], ..., z[\max(n-K, n_{max} - w_{sb})]\} < \xi + 0.5]T_c \qquad (16)$$

4. Simulation Results

The parameters of the TH-IR system considered here are as follows: the shape of the received pulses is the second derivative of the Gaussian pulse. Two models of indoor residential channels in LOS conditions, the CM1 model, and

the indoor office environment in LOS conditions, the CM3 model, have been used. The channel realizations are sampled at 10 GHz, and 100 different realizations are produced; each realization has one TOA. The criterion for making comparison between our method and other methods is the mean absolute error (MAE). The variance of the noise is assumed to be 1, and E_s is calculated for obtaining different $\frac{E_s}{N_0}$ ratios. To model the MUI, one interfering user is considered; that is U=2. The simulations were done for five situations: $\frac{E_s^{(2)}}{N_0} = 0, 10, 20, 30 dB$, and the situation in which there is no interfering user.

The other parameters of the simulations are $T_c = 4ns$, $T_f = 128ns$, $T_{sym} = 512ns$, $B = 500MHz$, $w_{sb} = 15$ (according to 60 ns). The number of the averaged symbols is $N_{sym} = 1000$. Three methods compared with each other are i) the old method with the use of $N_h = N_c/2$ and a constant TH code among different symbols (random within different frames), and the threshold given in Ref. [9], ii) the method i and using averaging nonlinear filtering [18], and iii) the new method with $N_h = N_c$, based on random TH code among different symbols, and the threshold obtained by our proposed method. The simulation results of channel CM1 shown in Figs. 2-6, and of channel CM3 shown in Figs. 7-11 indicate that the ranging error of the present method is less than that of the other methods, particularly in the presence of strong interferences and high values of $\frac{E_s}{N_0}$.
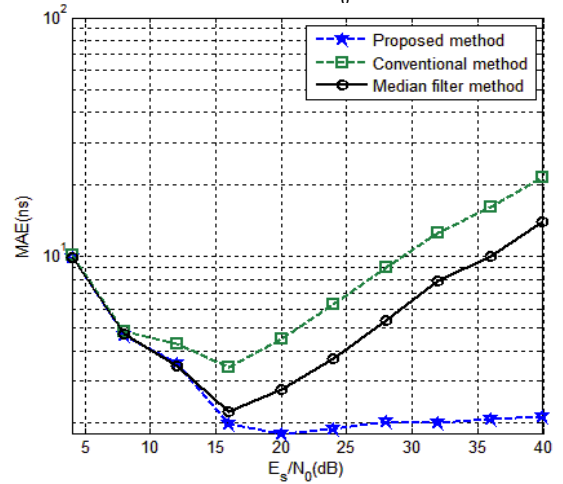
Fig. 2 MAE with no interference in CM1

In the presence of interference at levels of $\frac{E_s^{(2)}}{N_0} = 0, 10, 20 dB$ the proposed method can achieve the error below 4ns that is equivalent to a chip at high $\frac{E_s}{N_0}$ (12 dB or higher). Even when the interference is high, like at $\frac{E_s^{(2)}}{N_0} = 30 dB$, our method can reduce the estimation error at higher $\frac{E_s}{N_0}$, whereas in other compared methods not exist this decreased error.

Fig. 3 MAE with interference user s $\frac{E_S}{N_0} = 0dB$ in CM1

Fig. 4 MAE with interference user s $\frac{E_S}{N_0} = 10dB$ in CM1

Fig. 5 MAE with interference user s $\frac{E_S}{N_0} = 20dB$ in CM1

Fig. 6 MAE with interference user s $\frac{E_S}{N_0} = 30dB$ in CM1

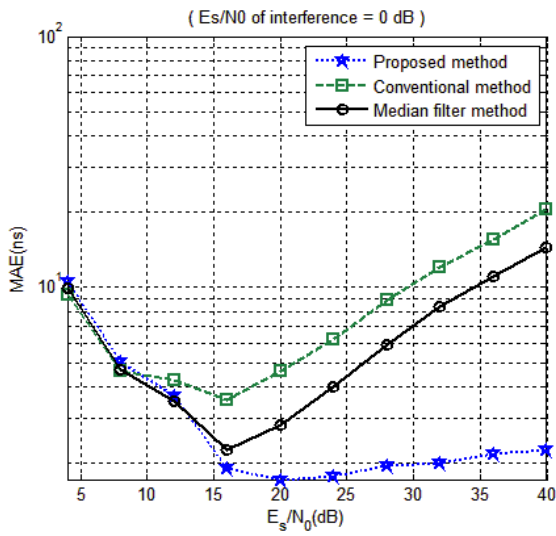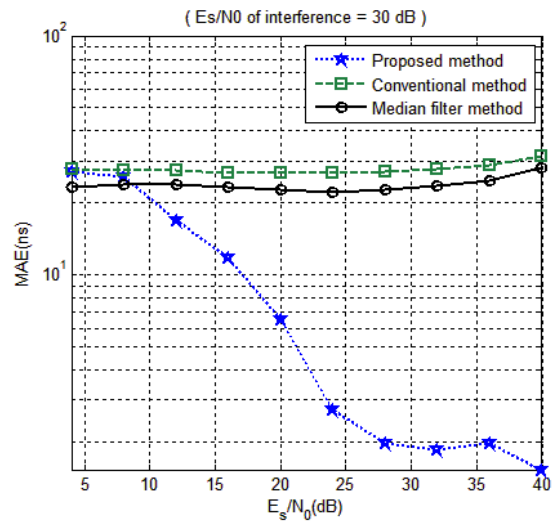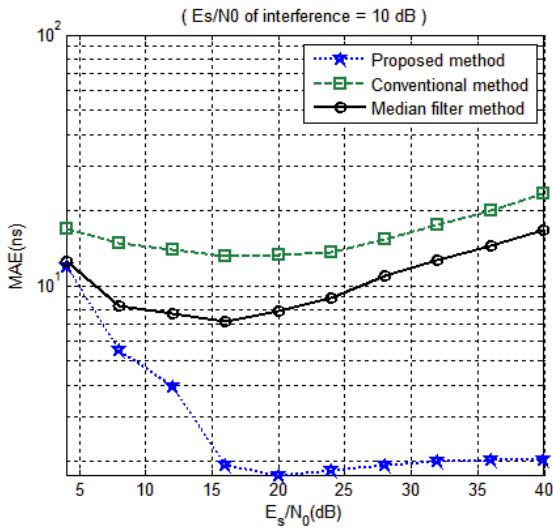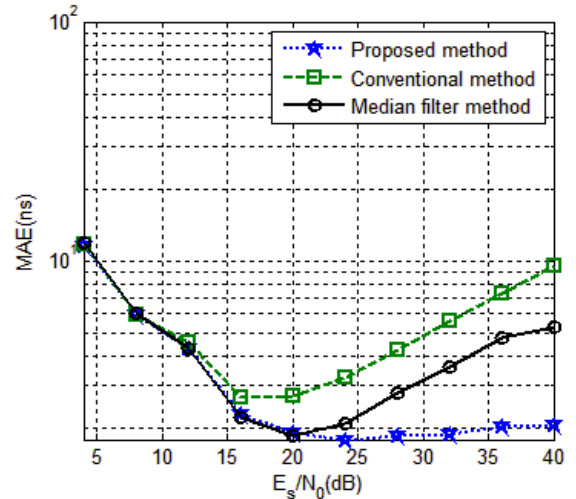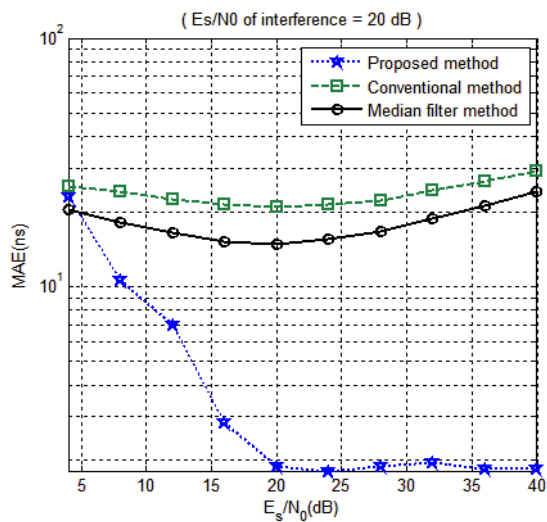Fig. 7 MAE with no interference in CM3

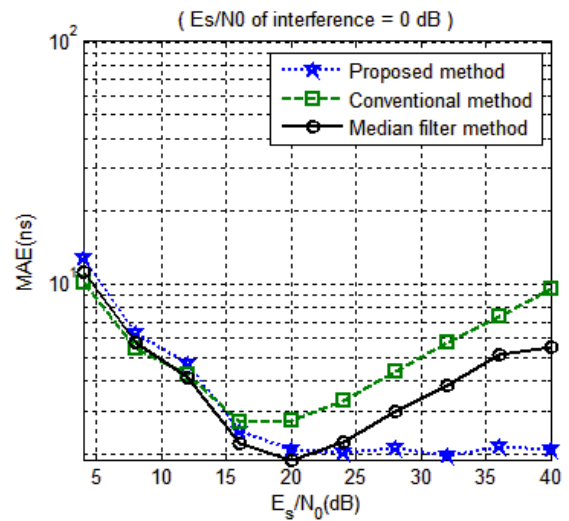Fig. 8 MAE with interference user s $\frac{E_S}{N_0} = 0dB$ in CM3

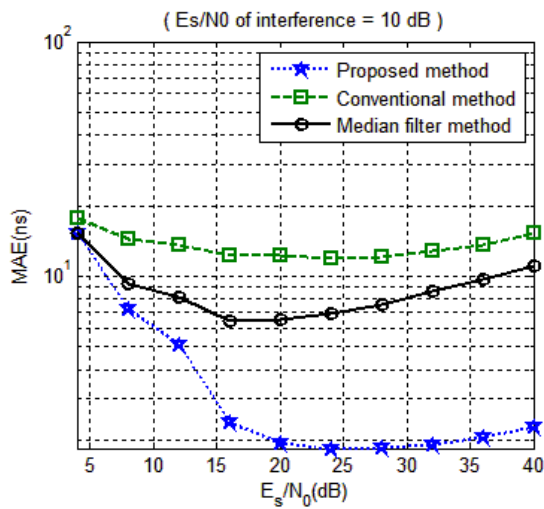Fig. 9 MAE with interference user□s $\frac{E_S}{N_0} = 10 dB$ in CM3

Fig. 11 MAE with interference user□s $\frac{E_S}{N_0} = 30 dB$ in CM3
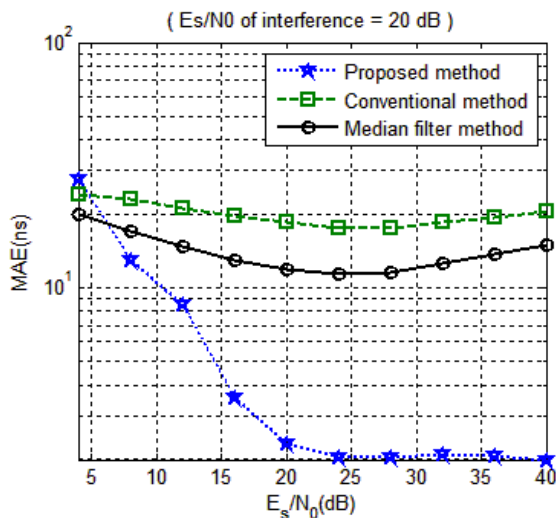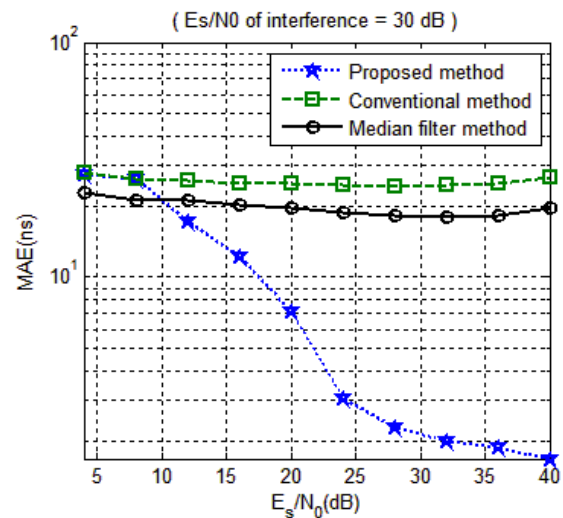
5. Conclusions

A new method based on the ED has been put forward for improving the estimation of the TOA in UWB systems with multi-path channels and in the presence of interference. The problem of the multi-path channels has been tackled with the use of the SBSMC algorithm. Moreover, a new approach, which can be called interference smoothing, has been suggested for lowering the destructive effect of the interference on the TOA estimation. This method can increase the accuracy of the distance measurements, particularly in the presence of strong interferences; indeed, it can remove the problem of the earlier-than-the-first-path detection which usually occurs due to the presence of interference. The capability of the present method has also been proven through some simulations.

Fig. 10 MAE with interference user□s $\frac{E_S}{N_0} = 20 dB$ in CM3

References

[1] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molish, H. V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios," *IEEE Signal Processing Mag.*, Vol. 22, No. 4, July 2005, pp. 70–84.

[2] D. Dardari, C.-C. Chong, and M. Z. Win, "Analysis of threshold-based TOA estimator in UWB channels," in Proc. Eur. Signal Process. Conf. (EUSIPCO), Florence, Italy, Sep. 2006.

[3] I. Guvenc and C. C. Chong , " A survey on TOA based wireless localization and NLOS mitigation techniques," *IEEE Commun. Surv. Tutorials*, Vol. 11, No. 3, Q3 2009, pp. 107 – 124.

[4] V. Lottici, A. D'Andrea, and U. Mengali, "Channel estimation for ultra-wideband communications," *IEEE Journal on Selected Areas in Communications*, Vol. 20, No. 9, 2002, pp. 1638–1645.

[5] M. Z. Win and R. A. Scholtz, "Characterization of ultrawide bandwidth wireless indoor channels: a communication theoretic view," *IEEE Journal on Selected Areas in Communications*, Vol. 20, No. 9, 2002, pp. 1613–1627.

[6] D. Dardari and M. Z. Win, "Threshold-based time-of-arrival estimators in UWB dense multipath channels," *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 10, Istanbul, Turkey, pp. 4723-4728, June 2006.

[7] I. Guvenc and Z. Sahinoglu, "TOA estimation with different IR-UWB transceiver types," *Proc. IEEE Int. Conf. UWB*, Zurich, Switzerland, Sep. 2005, pp. 426-431.

[8] I. Guvenc and Z. Sahinoglu, "Threshold-based TOA estimation for impulse radio UWB systems," *Proc. IEEE Int. Conf. UWB*, Zurich, Switzerland, Sep. 2005, pp. 420-425.

[9] I. Guvenc, Z. Sahinoglu, A. F. Molisch, and P. Orlik, "Non-coherent TOA estimation in IR-UWB systems with

different signal waveforms," *Proc. IEEE Int. Workshop on Ultrawideband Networks (UWBNETS)*, Boston, MA, October 2005, pp. 245-251.

[10] R. A. Scholtz and J. Y. Lee, "Problems in modeling UWB channels," *Proc. IEEE Asilomar Conf. Signals, Syst. Computers*, vol. 1, Pacific Grove, CA, Nov. 2002, pp. 706-711.

[11] I. Guvenc, and Z. Sahinoglu, "Threshold selection for UWB TOA estimation based on kurtosis analysis," *IEEE Commun. Lett.*, Vol. 9, No. 12, Dec. 2005, pp. 1025-1027.

[12] H. Zhang, X.R. Cui, and T. A. Gulliver, "Threshold Selection for UltraWideband TOA Estimation Based on Skewness Analysis," *Lecture Notes in Computer Science, Ubiquitous Intelligence and Computing, Springer*, Vol. 6905, Sept. 2011, pp. 503-513.

[13] H. Zhang, X.R. Cui and T. A. Gulliver, "Threshold Selection for TOA Estimation based on Skewness and Slope in Ultra-wideband Sensor Networks," *Journal of Networks*, Vol. 7, No. 7, July 2012, pp. 1038-1045.

[14] W. Y. Liu, H. Ding, X. T. Huang, and Z. L. Liu, "TOA estimation in IR UWB ranging with energy detection receiver using received signal characteristics," *IEEE Commun. Lett.*, Vol. 16, No. 5, 2012, pp. 738-741.

[15] W. Liu , H. Ding, X. Huang and Z. Liu, "Analysis of TOA estimation using the received signal characteristics in IR UWB ranging with energy detection receiver," *International Conference on Wireless Communications & Signal Processing (WCSP)*, Oct. 2012, pp. 1-6.

[16] H. Ding, W. Liu, X. Huang and L. Zheng, " First Path Detection Using Rank Test in IR UWB Ranging with Energy Detection Receiver under Harsh Environments," *IEEE Commun. Lett.*, Vol. 17, No. 4, April 2013, pp. 761-764.

[17] L. Yu, M. Laaraiedh, S. Avrillon, B. Uguen, J. Keignart, and J. Stephan, "Performance Evaluation of Threshold-Based TOA Estimation Techniques Using IR-UWB Indoor Measurements," in *Proc. of 18th European Wireless Conference (EW2012)*, Poznan, April 2012, pp. 1-7.

[18] T. Sakaguchi and T. ohtsuki, "A study on filterings in non-coherent TOA estimation in TH-UWB-IR systems," in *Proc. IEEE Conf. Commun.*, Oct. 2008, pp. 1-5.

[19] D. Dardari, A. Conti, U. Ferner, A. Giorgetti, and M. Z. Win, "Ranging with ultrawide bandwidth signals in multipath environments," *Proc. IEEE*, Vol. 97, No. 2, Feb. 2009, pp. 404-426.

[20] M. Ghasemlou, S. Nader-Esfahani, and V. Tabataba-Vakily, "A New Method of TOA Estimation in TH_UWB Systems with Multipath Channel and in the Presence of interference," 21st Iranian Conference on Electrical Engineering (ICEE), Mashhad, Iran, May 14-16, 2013, pp. 1-6 (in Persian).

[21] A. F. Molisch, K. Balakrishnan, D. Cassioli, C.-C. Chong, S. Emanmi, A. Fort, F. Karedal, J. Kunisch, H. Schantz, U. Schuster, and K. Siwiak, IEEE 802.15.4a Channel Model—Final Report, 2006, IEEE P802.15-04-0662-00-004a.

[22] D. Dardari, A. Giorgetti, and M. Z. Win, "Time-of arrival estimation of UWB signals in the presence of narrowband and wideband interference," *Proc. IEEE Int. Conf. on Ultra-Wideband*, Sept. 2007, pp. 71-76.

[23] Z. Sahinoglu and I. Guvenc, "Multiuser interference mitigation in noncoherent UWB ranging via nonlinear filtering," Proc. EURASIP Journal on Wireless Communication and Networking, Jan. 2006, pp.1–10.

[24] M. Ghasemlou, "Improvement of UWB Ranging Methods in Multipath Channels and in presence of Interference," M.Sc. Thesis, Faculty of Graduate studies, Department of Telecommunications engineering, Islamic Azad University Tehran South branch, Tehran, Iran, Feb. 2013.

Mahdieh Ghasemlou received the B.S. degree in Electrical engineering from Guilan University, Rasht, Iran, in 2008, and the M.Sc. degree in telecommunication engineering from Islamic Azad University Tehran South branch, Tehran, Iran, in 2013. Her research interests include the area of wireless communication systems, with emphasis on ultra wideband (UWB) communications, Ranging and localization techniques.

Saeid Nader-Esfahani received the B.S. degree in electrical engineering from University of Tehran, Tehran, Iran, in 1970, and the M.Sc. degree and the PHD degree in telecommunication engineering, University of Essex, England, in 1975 and 1978, respectively. He has done research in the areas of approximation and synthesis of filters, equalizations, adaptive filters, applications of cyclostationarity in signal processing, applications of HOS in signal processing, wireless communication systems. His current research interests include Blind Equalizations, OFDM systems, Ultra Wideband systems, compressive sensing, and cognitive radio. He is a Professor of Telecommunications of the University of Tehran. He has published more than 80 international journals, conference papers.

Vahid Tabataba-Vakili received the B.S. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 1970, and the M.Sc. degree and the PHD degree in communication systems engineering, Bradford University, England, in 1974 and 1978, respectively. His research interests include multi-user detection, space-time processing & coding, Bandwidth efficient digital modulation & coding, mobile radio cell planning, mobile cellular systems with emphasis on resource and mobility management, physical layer of wireless systems with emphasis on source/channel coding, MIMO and multi-access systems. He is a Professor of Telecommunications of Iran University of Science & Technology. He has published more than 110 international journals, conference papers.

Image Retrieval Using Color-Texture Features Extracted From Gabor-Walsh Wavelet Pyramid

Sajjad Mohammadzadeh*
Department of Electronics and Communications Eng., University of Birjand, Birjand, Iran
s.mohamadzadeh@birjand.ac.ir
Hasan Farsi
Department of Electronics and Communications Eng., University of Birjand, Birjand, Iran
hfarsi@birjand.ac.ir

Abstract

Image retrieval is one of the most applicable image processing techniques which have been extensively used. Feature extraction is one of the most important procedures used for interpretation and indexing images in Content-Based Image Retrieval (CBIR) systems. Effective storage, indexing and managing a large number of image collections are critical challenges in computer systems. There are many proposed methods to overcome these problems. However, the rate of image retrieval and speed of retrieval are still interesting fields of researches. In this paper, we propose a new method based on combination of Gabor filter and Walsh transform and Wavelet Pyramid (GWWP). The Crossover Point (CP) of precision and recall are considered as metrics to evaluate and compare different methods. The Obtained results show using GWWP provides better performance in compared to with other methods.

Keywords: Content-Based Image Retrieval (CBIR), Gabor Filter, Walsh Transform, Wavelet Pyramid Transform, Texture.

1. Introduction

Content Based Image Retrieval (CBIR) techniques are one of the most applicable and increasingly important topics in multimedia information systems [1]. An important building block in the image retrieval system is image indexing. Image Indexing is known as characterization of images based on some features of images [2]. Feature extraction is one of the most important procedures used for interpreting and indexing images in the CBIR systems [3]. There are many proposed methods and approaches for classification, indexing, searching and retrieval of visual information based on analysis of low-level image features, like color, texture and shape. [4].

Effective storage, transmission, indexing, managing a large number of image collection are serious challenges in computer systems [5]. Recently these challenges have been studied on different image databases and it has been attempted to solve these problems in computer vision [6] and image processing [7], [8]. The main goal of researchers in this field is to find a procedure to locate a desired image in a large and varied collection of image database. Traditional problems of image indexing methods such as taking a long time for manually indexing and huge required storage have led to rise of interest in retrieving images based on automatically derived features such as color, texture and shape which is known as CBIR [9], [10]. Nowadays CBIR technology is known as a form of commercial products such as QBIC [7] and Virage [8] in marketplace. However, because of some practical issues like, absence of hard evidence on the effectiveness of CBIR, this technology has not been used on the significant scale [8]. Many CBIR technology applications have been identified [7]. Medical, industrial and internet applications are some important examples of the applications. Nowadays color, texture or shape features are intensively used in image indexing. The combination of these features also showed more efficient performance in image retrieval [7].

In this paper we combine color and texture features and define the average of red, green and blue planes as gray plane. Then we extract some texture features from Gray plane. There are many different proposed methods to describe image texture. Texture analysis methods are divided into four categories: signal processing, model-based, geometrical and statistical introduced by Tuceryan and Jain [11]. We proposed only signal processing method for extracting texture feature. Texture feature is useless in image discrimination, if the variations of image intensity are highly uniform or non-uniform. The size of feature vectors and speed of retrieval are important aspects in performance of image retrieval. In this paper a novel method for image retrieval is proposed by using Gabor filter, Walsh transform and Wavelet Pyramid (GWWP) and then applied in database.

This paper has been structured as follows. In section 2 overview of the proposed CBIR system has been explained. Section 3 has been discussed parts of feature extraction proposed method. Haar transform has been explained in section 4. Similarity measurement is defined in section 5. Image retrieval using GWWP method has been described in section 6. In section 7, the obtained results have been presented.

---

* Corresponding Author

2. Content based image retrieval

Texture, color and shape features are basic features which used in CBIR systems. Texture and color features are absolutely easy for computing similarity. Some CBIR systems have combined texture and color feature to provide better performance and automatically retrieve relevant images from a large image database [12]. The standard CBIR system involves two important parts [12]. The first part extracts image features. This includes generating feature vectors of image in the database and representing the content of image accurately. The size of feature vectors must be extremely smaller than that of primary image. Similarity measurement is the second part of CBIR system. This part computes a distance between the query image and each image of database by using feature vectors of query image and each image in the database to obtain similar images.

Block diagram of the proposed CBIR system is shown in Figure 1. In this method, before extracting GWWP feature and generating feature vector, all images are resized to $256 \times 256 \times 3$ and gray plane with size of 256×256 is generated by averaging red, green and blue planes.

3. Feature extraction method

We have combined Gabor filter, Walsh transform and Wavelet Pyramid (GWWP) to extract texture features of image. The flowchart of GWWP method is represented in Figure 2 and the steps are represented as follows:

a. Applying Gabor filter (see section 3.1) on the gray plane with size of $N \times N$ separately. Gabor filter detects edges and lines of plane (image) and gives a maximum and suitable response at the edge. Orientation and frequency representations of Gabor filters are similar to those of the human visual system, and they have been found to be particularly appropriate for texture representation and discrimination [13].

b. Performing wavelet transform with size of $N \times N$ on the filter output with size of $N \times N$ to generate approximation (Low-Low), horizontal (Low-High), vertical (High-Low) and diagonal (High-High) components which are explained in section 3.2. We used approximation component for next step.

c. To construct modified approximation component by applying Walsh transform (see section 3.3) on the approximation component in step 'b'. Walsh transform can be reduced to subtraction and addition operations (no division or multiplication). This allows the use of simpler hardware and low complexity to calculate the transform and increases the speed of retrieval.

d. Applying inverse wavelet transform with modified approximation component and zeroing horizontal, vertical and diagonal components. This method is wavelet pyramid transform which is completely explained in section 3.2. New image is constructed by using inverse wavelet whereas some information are lost due to removing horizontal and vertical components. In next level, we need to new image to construct the new approximation and diagonal components.

e. To take alternative rows and columns by down-sampling the output in step 'd' with size of $N/2 \times N/2$. Down-sampling reduces the size of feature vector which is very important for increasing the speed of retrieval.

f. To construct GWWP of level-p by repeating steps 'b' to 'e', 'p' times on the each plane.

We consider approximation component level-p in step 'c' as GWWP feature of each image and store this feature as feature vector of image.

3.1 Gabor filter

Gabor filter or Gabor wavelet is a method used to extract features of image through analysis of the frequency domain rather than the spatial domain [13]. Eq. (1) represents the Gabor filter, where x and y shows the position of pixel in the spatial domain, ω_0 is radial center frequency, θ shows the orientation of the Gabor direction and σ is defined as the standard deviation of the Gaussian function along with the x and y axes where $\sigma_x = \sigma_y = \sigma$ [13]:

$$\psi(x, y, \omega_0, \theta) = \frac{1}{2\pi\sigma^2} exp\{-((xcos\theta + ysin\theta)^2 + (-xsin\theta + ycos\theta)^2)/2\sigma^2\} \times [exp\{i(\omega_0 xcos\theta + \omega_0 ysin\theta)\} - exp\{-\omega_0^2\sigma^2/2\}] \quad (1)$$

According to Eq. 1 Gabor filter can be decomposed into real and imaginary parts, which are given by:

$$\psi_r(x, y, \omega_0, \theta) = \frac{1}{2\pi\sigma^2} exp\left\{-\left(\frac{x'^2+y'^2}{\sigma^2}\right)\right\} \times \left[cos\omega_0 x' - e^{-\omega_0^2\sigma^2/2}\right] \quad (2)$$

$$\psi_i(x, y, \omega_0, \theta) = \frac{1}{2\pi\sigma^2} exp\left\{-\left(\frac{x'^2+y'^2}{\sigma^2}\right)\right\} \times sin\ \omega_0 x' \quad (3)$$

Where:

$$x' = x\ cos\theta + y\ sin\theta, \quad y' = -x\ sin\theta + y\ cos\theta$$

We consider $\sigma = \pi/\omega_0$. Gabor features, $C_{\Psi I}$, can be achieved by using the convolution of image, I, and Gabor filter, Ψ, as represented in Eq. 4 [13]:

$$C_{\psi 1} = I(x, y) * \psi(x, y, \omega_0, \theta) \quad (4)$$

As mentioned, Gabor filter can be written as the summation of real and imaginary parts shown in Eq. 2 and Eq. 3. Therefore we can compute real and imaginary parts of $\psi(x, y, \omega_0, \theta)$, represented by $C_{\Psi I}^r$ and $C_{\Psi I}^i$ respectively, by replacing Eq. 2 and Eq. 3 in the term of $\psi(x, y, \omega_0, \theta)$ in Eq. 4. Local properties of the image can be achieved using real and imaginary parts, which is given by:

$$C_{\Psi I}\left(x, y, \omega_0, \theta\right) = \sqrt{\left\|C_{\Psi I}^r\right\|^2 + \left\|C_{\Psi I}^i\right\|^2}. \quad (5)$$

A fast convolution method is utilized using scanning windows by applying a one-time convolution with Fast Fourier Transform (FFT), point-to-point multiplication and Inverse Fast Fourier Transform (IFFT).The values of radial

center frequencies and orientations used in this paper are shown in Eq. 6, where n ∈ {0, 1, 2} and m ∈ {0, 1, 2, …, 7} [13]:

$$\omega_n = \frac{\pi}{2\sqrt{2}^n} \, , \qquad \theta_m = \frac{\pi}{8} m \qquad (6)$$

3.2 Wavelet pyramid transform

Wavelet transform is multi-level signal decomposition. It represents a signal as a basis function superposition called wavelets [14]. Wavelet transform has several interesting properties [14]. The principal property of the proposed wavelet feature analyzes the signal in at various frequency bands giving higher frequency resolution and lower time resolution at lower frequencies, lower frequency resolution and higher time resolution at higher frequencies which have been shown in Figure 3 [14]. For a given image with size of $N \times N$, the two-dimensional Haar wavelet transform includes log_2^N stages. The first stage provides four sets of coefficients known as, approximation coefficients cA_1, horizontal coefficients cH_1, vertical coefficients cV_1, and diagonal coefficients cD_1. These sets are computed by convolving columns or rows of image with the low-pass filter for approximation, and with the high-pass filter, which are followed by dyadic decimation (down-sampling). The length of these filters is 2n sample. Therefore, if the length of an image is N, then the length of output signal using low-pass and high-pass filters will be $N + 2n - 1$ [15], [16]. Figure 4 describes a flowchart of the basic decomposition of wavelet transform for an input image. The next step splits the approximation coefficients cA_1, in two separate parts using the same method, described above, replacing input image (s) by cA_1, and producing cA_2, cH_2, cV_2 and cD_2, and so on [17]. As an example, three levels of wavelet pyramid is shown in Figure 5.

3.2.1 Haar transform

Haar transform was proposed in 1909 by Alfréd Haar [20]. Haar used this transform to give an example of a countable orthonormal system for the space of square-Integrable functions on the real line [20]. The Haar wavelet is also one of the simplest possible wavelet. The discrete entity of Haar wavelet transform is one of the technical disadvantages of this method. However, this property can be considered as an advantage for analysis of signals with sudden transition like monitoring of tool failure in machines [21].

The Haar mother wavelet function $\psi(t)$ can be described as:

$$\psi(t) = \begin{cases} 1 & , 0 \le t \le \frac{1}{2} \\ -1 & , \frac{1}{2} \le t \le 1 \\ 0 & , otherwise \end{cases} \qquad (7)$$

And its scaling function $\varphi(t)$ is given as:

$$\varphi(t) = \begin{cases} 1 & , 0 \le t \le 1 \\ 0 & , otherwise \end{cases} \qquad (8)$$

3.3 Walsh transform

Walsh transform matrix [18] is defined as a set of N rows, denoted W_j, for $j = 0, 1, …, N - 1$. The properties of Walsh transform matrix are described as:

a. W_j takes on the values + 1 and - 1.
b. $W_j[0] = 1$ for all j.
c. $W_j \times W_k^T = 0$, for $j \ne k$ and $W_j \times W_k^T = N$, for $j = k$.
d. W_j has exactly j zero crossings, for $j = 0, 1, …, N - 1$.

Each row W_j is either even or odd with respect to its midpoint.

Hadamard matrix of order N is used to define Walsh transform matrix. The row of Walsh transform matrix is the row of the Hadamard matrix determined by the Walsh code index, which is an integer in the range $[0, …, N - 1]$. For the Walsh code index equal to an integer j, the respective Hadamard output code has exactly j zero crossings, for $j = 0, 1, …, N - 1$.

Following stages show the Kekre's Algorithm to generate Walsh Transform from Hadamard matrix [19]:

a. The N coefficients of Walsh matrix are arranged in a row and then the row is split to segments with length of $N/2$, one segment in forward order and the other part is written in reverse order:

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \\ 15 & 14 & 13 & 12 & 11 & 10 & 9 & 8 & & & & & & & & \end{bmatrix}$$

Now we have two rows, each of these rows are again split in $N/2$ parts and other part is written in reverse order below the upper rows as:

$$\begin{bmatrix} 0 & 1 & 2 & 3 \\ 15 & 14 & 13 & 12 \\ 7 & 6 & 5 & 4 \\ 8 & 9 & 10 & 11 \end{bmatrix}$$

This step continues, until a single column giving the order of Hadamard rows, is achieved. The result sequence is presented in following:

$$[0, 15, 7, 8, 3, 12, 4, 11, 1, 14, 6, 9, 2, 13, 5, 10]$$

b. Based on this sequence the Hadamard rows are arranged to generate Walsh transform matrix. The Walsh transform of the given image is calculated by product of Walsh matrix and image matrix.

The number of additions required to apply Walsh transform on an image with size of $N \times N$ are $2N^2 \times (N - 1)$ [19].

4. Similarity measurements

Traditional Euclidean Distance (ED) is the most common metric used to compute match or similarity value in CBIR system to obtain relevant images.

If Feature Vector of Database (FVD) and Feature Vector of Query (FVQ) are two-dimensional feature vectors of database image and query image respectively, then the Euclidean distance of these feature vectors is obtained by:

$$ED = \sqrt{\sum_{i=1}^{n}\left(FVD_i - FVQ_i\right)^2} \qquad (9)$$

5. Image retrieval using GWWP method

5.1 Feature extraction

We generate feature vectors for image database by applying GWWP level-1, level-2, …, level-7 and store approximation components as feature vectors for each image. The size of feature vector is $N/2 \times N/2$ in GWWP level-1 and the size of feature vector in GWWP level-2 is $N/4 \times N/4$. Gray image is defined as the average of red, green and blue planes components and is used to generate Gray-GWWPs feature vector to obtain respective images for different levels.

5.2 Query execution

We use the proposed method to extract query feature and generate feature vector of query image. Then relevant images are retrieved by comparing feature vector of query with feature vectors of database in level-p using Euclidian distance as similarity measure.

The proposed method reduces the size of feature vector and computation time extremely in high level of GWWP and gives better precision and recall values. Complete GWWP needs $2N^2 \times (N-1)$ additions and GWWP of level-p needs $2(N/_2 P)^2 \times ((N/_2 P) - 1)$ addition for image with size of $N \times N$.

6. Experimental Results

We have compared performance of the proposed method with precision and recall criteria. These two standard criteria are given by:

$$Precision = \frac{Number\_of\_Relevant\_Images\_Retrieved}{Total\_Number\_of\_Images\_Retrieved} \qquad (10)$$

$$Recall = \frac{Number\_of\_Relevant\_Images\_Retrieved}{Total\_Number\_of\_Relevant\_Images\_in\_Database} \qquad (11)$$

In order to evaluate the proposed method in image retrieval, we first provide an image database [22] including 1000 variable size images. These images are classified in 11 classes of human being, horse, elephant, flower, bus, manmade thing and natural scenery. We have tried to collect a diverse database set to evaluate the proposed method. Figure 6 shows sample images from the database. After collecting a proper image database, we select 5 images from each class (55 random images) as query images to evaluate the proposed system with other methods. As described in previous sections, the features of each query image is extracted using GWWP method, and feature vectors of query images are constructed. Then Euclidian distance between each query feature vector of image and database feature vectors are computed by using Eq. 9. The obtained distances are sorted based on which images have minimum distance

with query image to find the best relevant images in database. Then, the number of relevant images is computed and the precision and recall for each number of retrieved images for all 55 query images are obtained by using Eq. 10 and Eq. 11. We next consider the average of these 55 precisions and recalls for each number of retrieved images as the precision and recall for each method.

For example in Figure 7 sixteen closest images for the sample query image have represented by using the GWWP level-5 and computing Euclidian distances between query image and all images in database and sorting distances based on minimum distance. The average precision and recall for Haar wavelet, Walsh wavelet and GWWP methods (the proposed method) are obtained and plotted by grouping the number of retrieved images in Figures 8, 9 and 10, respectively. Note that Haar wavelet method has not used Walsh matrix and Gabor filter and Walsh wavelet has not utilized Haar transform and Gabor filter. According to these Figures it is obvious that the precision decreases and the recall increases by increasing the number of retrieved images. Also by increasing wavelet level until level-5, the precision and the recall increase but they decrease after level-5. For example the precision of GWWP level-1, 3 and 5 are 0.358, 0.366 and 0.382 in 90 of retrieved image respectively but the precision of GWWP level-7 is 0.337. Therefore precision and recall level-5 of wavelet methods are better than other wavelet levels. Moreover it is observed that precision of Haar wavelet and Walsh wavelet level-5 are 0.338 and 0.344 respectively, whereas the precision of GWWP level-5 is 0.384 in 88 of retrieved image. Based on these figures the proposed GWWP method provides better performance rather than Haar wavelet and Walsh wavelet.

The other important parameter for comparing CBIR techniques is the percent of CP of precision and recall [11] which the CP occurred in 91 of retrieved image in our database because the most classes have 91 images that is specified in Figure 8, 9 and 10 clearly. The percent of CP for different levels of Haar wavelet, Walsh wavelet, GWWP, dominant color descriptor (DCD) [23], scalable color descriptor (SCD) [23] methods and the size of feature vector have been represented in Table 1. As observed in Table 1, the size of feature vector decreases by increasing wavelet level but the CP only increases until level-5. However the size of feature vector is 4 in level-7 but the CP (31.21% of CP of GWWP) is less than level-1 with 16384 size of feature vector which CP of GWWP level-1 is 35.81%. It is obvious that level-5 of each three wavelet methods with 64 size of feature vector has higher CP of precision and recall than other levels in each method. In Table 1, it could be observed that the GWWP level-5 with CP of 38.84% and size of 64 feature vectors provides the best performance compared to other methods. However, SCD method with CP of 37.05% is nearest to GWWP level-5, but size of feature vector in SCD method is 121 which is 1.9 times of feature vectors of GWWP level-5. The size of feature vector is very important parameter for increasing speed of computing distance and retrieval and

for decreasing storage space of feature vector. However, as shown in Table 1, the proposed GWWP method in level-1 (CP = 35.81%) has even better performance than Walsh wavelet (CP = 34.03%) and Haar wavelet (CP = 33.60%) in level-5. Therefore GWWP technique can be considered as a more powerful method than Haar, Walsh wavelet, DCD and SCD. Moreover, the proposed system not only reduces the size of feature vector and storage space but also improves the performance of image retrieval.

7. Conclusions

Image retrieval is an applicable technique finds relevant images in a large image database. By increasing the size of database, the challenge of average reduction in precision and recall and speed of retrieval become more serious. In this paper, we proposed a new method based on Gabor filter, Walsh transform and wavelet pyramid to improve the performance of image retrieval. We used an average of three planes of red, green and blue planes to generate gray plane to examine the proposed method. The obtained results show using Gabor filter before Walsh transform and wavelet pyramid can improve the performance of image retrieval in our database. Specially, wavelet in level-5 results in better performance in compared to other levels of wavelet pyramid. Moreover, the GWWP level-5 reduces the size of feature vectors and storage space and provides higher performance for image retrieval at the same time.

Fig. 1. Block diagram of CBIR system.
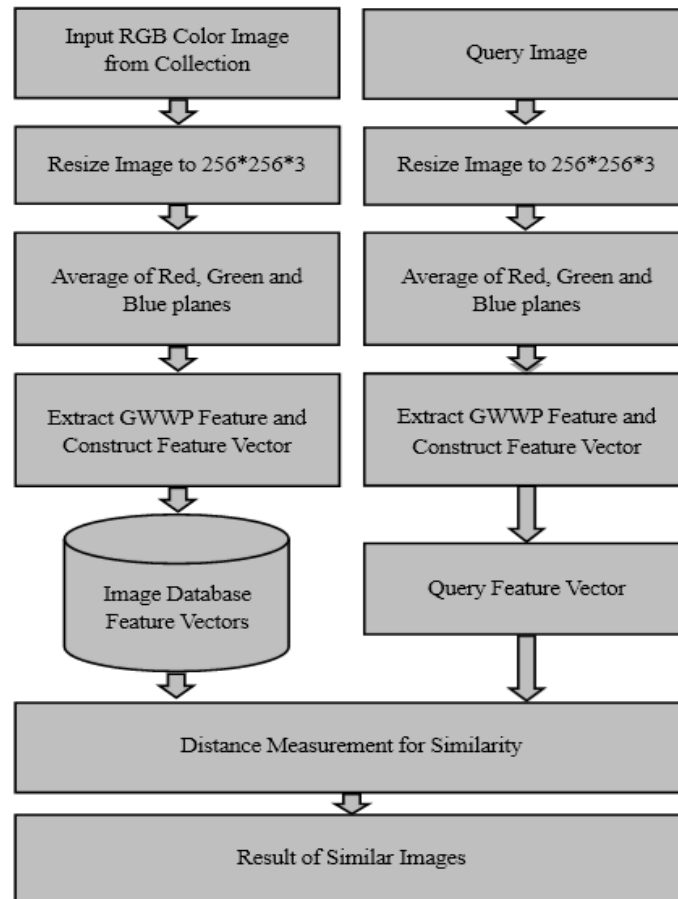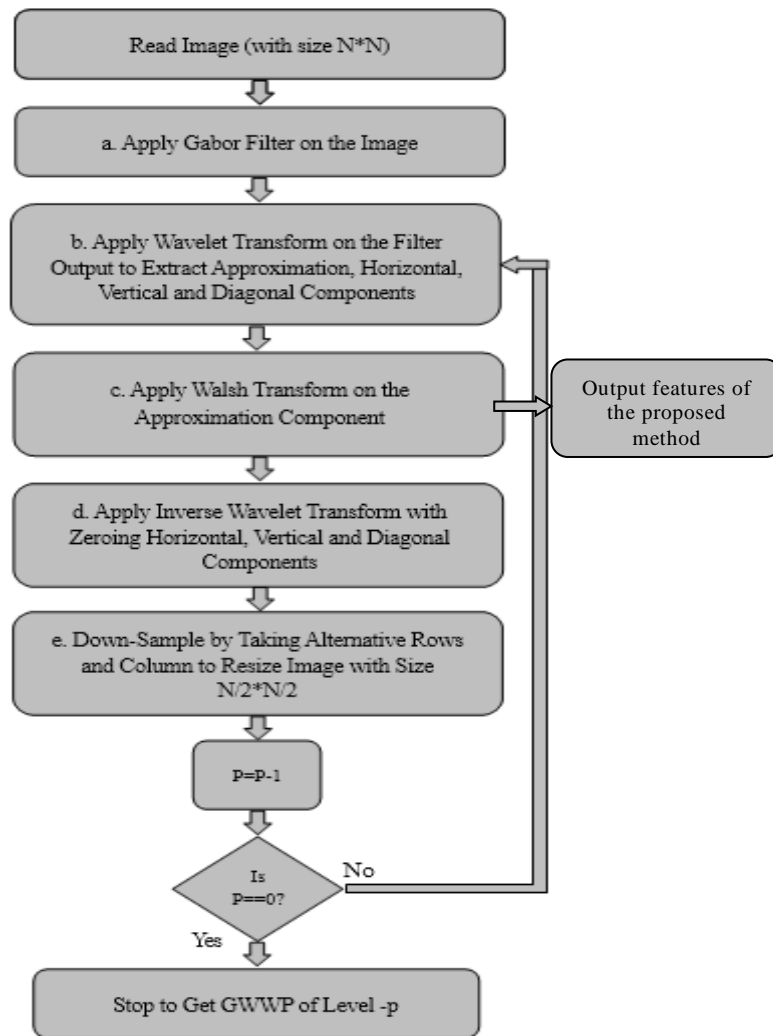
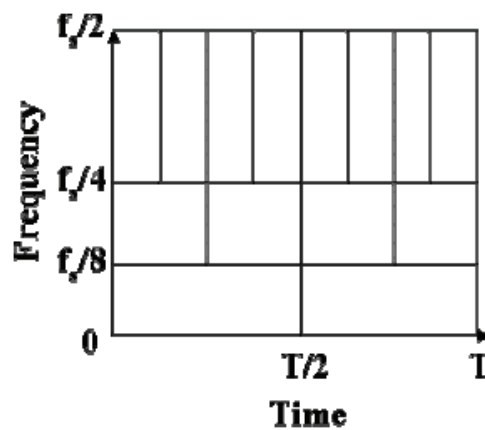Fig. 2. Flowchart for generating GWWP of level-p.

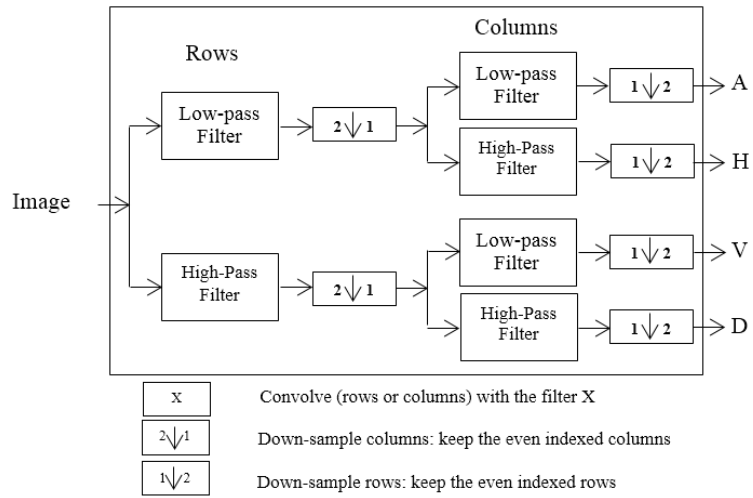Fig. 3. Time frequency resolution of wavelet transform [14]

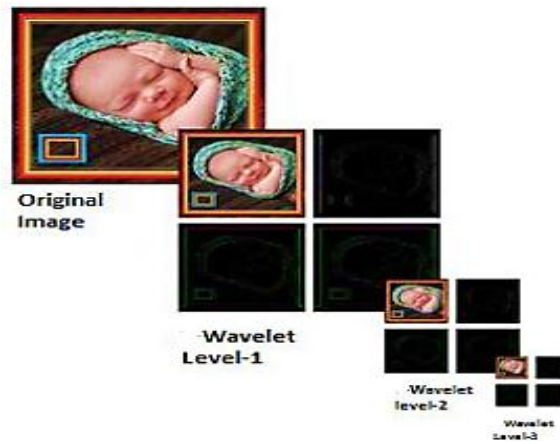Fig. 4. Decomposition of the image or cA.

Fig. 5. Different levels of wavelet pyramid [12].
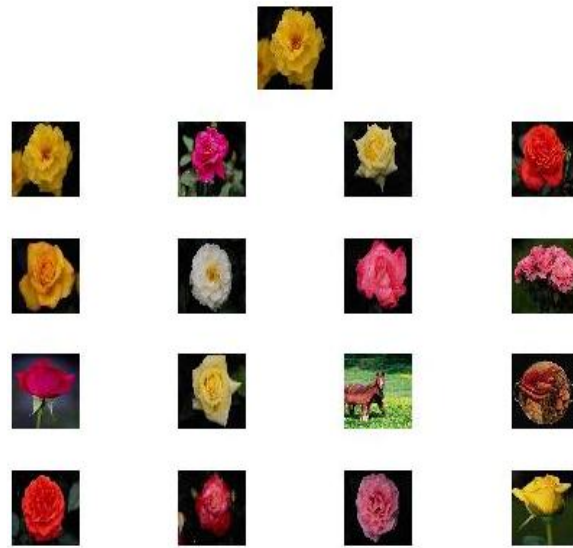

Fig. 6. Sample images of database.

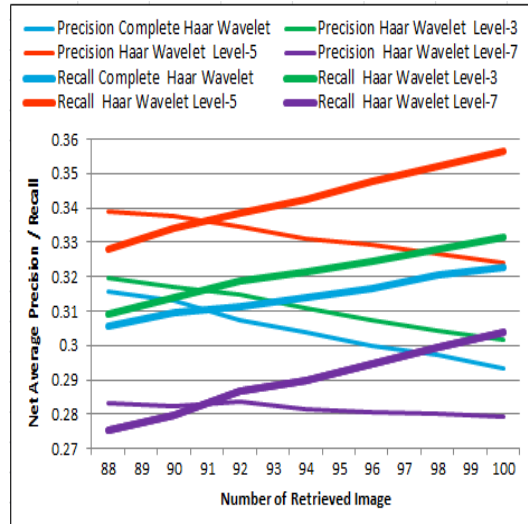Fig. 7. Image retrieved using GWWP level-5.
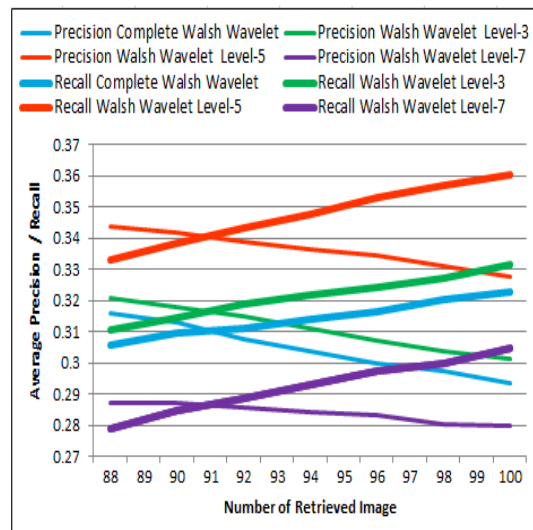

Fig. 8. Haar wavelet precision\ recall.

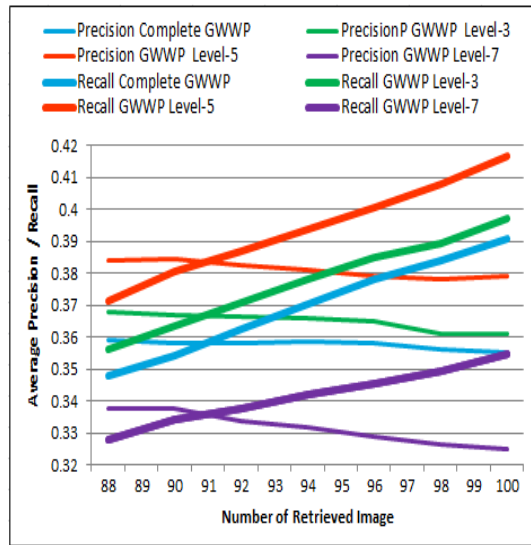Fig. 9. Walsh wavelet precision\ recall.

Fig. 10. Gray-GWWP method precision\ recall.

Table 1. Size of feature vector and the present of CP of different methods

Type			Size of feature vector	Crossover point
Haar wavelet		Level 1	16384	31.02%
		Level 3	1024	31.60%
		Level 5	64	33.60%
		Level 7	4	28.33%
Walsh wavelet		Level 1	16384	31.02%
		Level 3	1024	31.64%
		Level 5	64	34.03%
		Level 7	4	28.64%
GWWP wavelet		Level 1	16384	35.81%
		Level 3	1024	36.67%
		Level 5	64	38.34%
		Level 7	4	31.21%
DCD			32	36.37%
SCD			121	37.05%

References

[1] C. H. Lin, C. W. Liu, H. Y. Chen, "Image Retrieval and Classification Using Adaptive Local Binary Patterns Based on Texture Features", IET Image Processing, Vol. 6, No. 7, 2012, pp. 822 - 830.

[2] A. Marakakis, G. Siolas, N. Galatsanos, A. Likas, A. Stafylopatis, "Relevance Feedback Approach for Image Retrieval Combining Support Vector Machines and Adapted Gaussian Mixture Models', IET Image Processing, Vol. 5, No. 6, 2011, pp. 531-574.

[3] O. Starostenko, A. Chávez-Aragón, G. Burlak, R. Contreras, "A Novel Star Field Approach for Shape Indexing in CBIR System", J. of Eng. Letters, Vol. 1, 2007, pp. 10-21.

[4] Q. Zhang, E. Izquierdo, "Histology Image Retrieval in Optimized Multi-feature Spaces" IEEE Journal of Biomedical and Health Informatics, Vol. 7 , No. 1, 2013, pp. 240 - 249

[5] H. B. Kekre, S. D. Thepade, "Rendering Futuristic Image Retrieval System", Proc. EC2IT, 2009.

[6] B. G. Prasad, K. K. Biswas, S. K. Gupta, "Region–Based Image Retrieval Using Integrated Color, Shape, and Location Index", International Journal on Computer Vision and Image Understanding Special Issue: Colour for Image Indexing and Retrieval, Vol. 94, 2004, pp. 193-233.

[7] R. Montagna, G. D. Finlayson, "Padua Point interpolation and Lp-Norm Minimization Color-based Image Indexing and Retrieval", IET Image Processing, Vol. 6, 2012, pp. 139-147.

[8] D. N. Vizireanu, S. Halunga, G. Marghescu, "Morphological Skeleton Decomposition Inter-Frame Interpolation Method", Journal of Electronic Imaging, Vol. 19, No. 2, 2012, pp. 1-3.

[9] H. B. Kekre, T. K. Sarode, S. D. Thepade, "Image Retrieval Using Color-Texture Features from DCT on VQ Code Vectors Obtained by Kekre's Fast Codebook Generation", ICGST International Journal on Graphics, Vision and Image Processing (GVIP), Vol. 9, 2009, pp. 1-8.

[10] H. B. Kekre, T. K. Sarode, S. D. Thepade, "Color-Texture Feature Based Image Retrieval Using DCT Applied on Kekre's Median Codebook", International Journal on Imaging (IJI), Vol. 2, 2009, pp. 55-65.

[11] E. Ozdemir, G. Demir, C. Browse, "A Hybrid Classification Model for Digital Pathology Using Structural and Statistical Pattern Recognition", IEEE Transactions on Medical Imaging, Vol. 32, No.2, 2013, pp. 474 – 483

[12] H. B. Kekre, S. D. Thepade, "Using YUV Color Space to Hoist the Performance of Block Truncation Coding for Image Retrieval", Proc. IEEE-IACC'09, 2009.

[13] W. H. Yap, M. Khalid, R. Yusof, "Face Verification with Gabor Representation and Support Vector Machines", Proc. IEEE-AMS '07, 2007, p. 451.

[14] A. Lakshmi, S. Rakshit, "New Wavelet Features for Image Indexing and Retrieval", IEEE 2nd International Advance Computing Conf., 2010, pp. 145-150.

[15] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, C. Roux, "Fast Wavelet-Based Image Characterization for Highly Adaptive Image Retrieval", IEEE Transactions on Image Processing, Vol. 21, No. 4, 2012, pp. 1613 – 1623.

[16] A. Lakshmi, S. Rakshit, "New Wavelet Features for Image Indexing and Retrieval", IEEE 2nd International Advance Computing Conf., 2010, pp. 145-150.

[17] Yu Pan, Li Chai, Yuxia Sheng, "Computation and Optimization of Frame Bounds for the Laplacian Pyramid", 25th Chines Conference Control and Decision, 2013, pp. 1423 - 1428

[18] H. B. Kekre, S. D. Thepade, A. Maloo, "Performance Comparison of Image Retrieval Techniques Using Wavelet Pyramids of Walsh, Haar and Kekre Transforms", International Journal of Computer Applications, Vol. 4, 2010, pp. 1-8.

[19] H. B. Kekre, V. Bharadi, "Walsh Coefficients of the Horizontal and Vertical Pixel Distribution of Signature Template", Proc. ICIP '07, 2007, p. 10.

[20] S. G. Sathyanarayana, A. Gargava, S. M. Venkatesan, "Parameterized Transform Domain Computation of the Hilbert Transform Applied to Separation of Channels in Doppler Spectra," IEEE 3rd International Advance Computing Conference (IACC), 2013, pp. 1189 - 1194

[21] A. M. Atto, Y. Berthoumieu, P. Bolon, "2-D Wavelet Packet Spectrum for Texture Analysis IEEE Transactions on Image Processing, Vol. 22, No. 6, 2013, pp. 2495 - 2500

[22] http://wang.ist.psu.edu/docs/related/Image.orig (last referred on June, 10th, 2009).

[23] Q. Jiang, Weina We, H. Zhang, "New Researches About Dominant Color Descriptor and Graph Edit Distance', Int. Conf. Intell. Human-Mach. Syst. and Cybernetics (IHMSC), Vol. 1, 2011, pp. 50-52

Sajjad Mohammadzadeh received the B.Sc. degree in electrical engineering from Sistan & Baloochestan, University of Zahedan, Iran, in 2010. He received the M.Sc. degree in communication engineering from South of Khorasan, University of Birjand, Birjand, Iran, in 2012. He is currently Ph. D student in Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran. His area research interests include Image Processing and retrieval, Pattern recognition, Digital Signal Processing and Sparse representation. His email address is: s.mohamadzadeh@birjand.ac.ir.

Hasan Farsi received the B.Sc. and M.Sc. degrees from Sharif University of Technology, Tehran, Iran, in 1992 and 1995, respectively. Since 2000, he started his Ph.D in the Centre of Communications Systems Research (CCSR), University of Surrey, Guildford, UK, and received the Ph.D degree in 2004. He is interested in speech, image and video processing on wireless communications. Now, he works as associate professor in communication engineering in department of Electrical and Computer Eng., university of Birjand, Birjand, IRAN. His Email is: hfarsi@birjand.ac.ir.

Language Model Adaptation Using Dirichlet Class Language Model Based on Part-of-Speech

Ali Hatami*

Computer Engineering Department, Iran University of Science and Technology, Tehran, Iran

ali_hatami@comp.iust.ac.ir

Ahmad Akbari

Computer Engineering Department, Iran University of Science and Technology, Tehran, Iran

akbari@iust.ac.ir

Babak Nasersharif

Electrical and Computer Engineering Department, K. N. Toosi University of Technology, Tehran, Iran

bnasersharif@kntu.ac.ir

Abstract

Language modeling has many applications in a large variety of domains. Performance of this model depends on its adaptation to a particular style of data. Accordingly, adaptation methods endeavour to apply syntactic and semantic characteristics of the language for language modeling. The previous adaptation methods such as family of Dirichlet class language model (DCLM) extract class of history words. These methods due to lake of syntactic information are not suitable for high morphology languages such as Farsi. In this paper, we present an idea for using syntactic information such as part-of-speech (POS) in DCLM for combining with one of the language models of n-gram family. In our work, word clustering is based on POS of previous words and history words in DCLM. The performance of language models are evaluated on BijanKhan corpus using a hidden Markov model based ASR system. The results show that use of POS information along with history words and class of history words improves performance of language model, and decreases the perplexity on our corpus. Exploiting POS information along with DCLM, the word error rate of the ASR system decreases by 1.2% compared to DCLM.

Keywords: Speech Recognition, Language Model Adaptation, Part-of-Speech, Perplexity, Word Error Rate.

1. Introduction

Statistical language modeling (SLM) has been successfully applied to many natural language and speech processing. The purpose of LM is to assign probabilities to sequences of words according to a certain distribution. Speech recognition focuses on searching for the best word sequence $\widehat{W}$ by maximizing a posteriori (MAP) probability of speech utterance X [1]:

$$\widehat{W} = \text{argmax } p(W|X) = \text{argmax } p(X|W)\, p(W) \qquad (1)$$

where $p(X|W)$ is the acoustic likelihood given the hidden Markov model (HMM), and $p(W)$ is the prior word probability given the LM. N-gram LM is a known approach that assigns probability to next word based on its immediately preceding n-1 history words. In an n-gram model [2], the probability of a word sequence $(w_1^T = (w_1, \dots, w_T))$ is calculated by multiplying the probabilities of predicted word w_i conditioned on its preceding $n-1$ words depicted by w_{i-n+1}^{i-1}:

$$p(W) = \prod_{i=1}^{T} p\big(w_i|w_1^{i-1}\big) \cong \prod_{i=1}^{T} p\big(w_i|w_{i-n+1}^{i-1}\big) \qquad (2)$$

where $p\big(w_i|w_1^{i-1}\big)$ shows the conditional probability of w_i given w_1^{i-1}. Factored language model (FLM) [3] is another kind of n-gram models. The FLM was proposed using factors for each word. In a FLM, a word is considered as a vector of K factors.

$$w_t = \{f_t^1, \dots, f_t^K\} \qquad (3)$$

These factors can be anything, including morphological classes, stems, roots and other such features.

The n-gram models suffer from the insufficiencies of long-distance information, which limit the model performance. To compensate this, n-gram model can be combined with the adaptation methods like latent Dirichlet allocation (LDA) that extract the semantic information. LDA [4] provides a powerful mechanism for discovering the structure of a text document. The latent topic of each document is treated as a random variable. To tackle the data sparseness and extract the large-span information for n-gram models, in [5], a new Dirichlet class LM (DCLM) is constructed. In this technique, the latent variable reflects the class of an n-gram event rather than the topic in LDA model. In addition, Cache DCLM (CDCLM) [5] is proposed to improve DCLM by considering dynamic classes of history words in the online estimation.

The previous adaptation methods just used semantic information and did not consider syntactic features. In the languages with high morphology such as Farsi, exploiting the syntactic information such as POS can be useful.

In this paper, we proposed a technique for using POS along with adaptation methods to improve language

---

* Corresponding Author

model and so speech recognition rate. In our DCLM based approach, word clustering is performed exploiting previous POS and history words. In this technique, we calculate the word probability given POS of previous words along with history words.

The remainder of this paper is organized as follows. In Section 2, we provide an overview of related works. In Section 3, we discuss the proposed technique for using POS along with adaptation methods. Section 4 evaluates the LMs performance. Finally, in Section 5 we conclude our paper.

2. Related Work

2.1 Latent Dirichlet Allocation

Topic-based model is a common method for extracting semantic information from text corpus in order to adapt a language model. In the last decade, a variety of probability topic modeling approaches has been proposed to analyze the latent topics and meaning of documents and words, such as latent Dirichlet allocation (LDA). Blei et al. [4] introduced LDA by incorporating the Dirichlet priors for extracting the topic structure of a document. LDA builds a hierarchical Bayesian model. In this model, documents are represents by the random latent topics, which are specified by the distributions over words. In other words, LDA discovered the topic at document level and were used for building topic-based language model [6].

LDA was shown effective in document classification [4] and speech recognition [5]. LDA model defines two parameters consist of { α, β }, where α denotes the Dirichlet parameters of topic z and β is a matrix that contains value of the topic unigram $\beta_{w,z} = p(w|z)$. A topic mixture vector θ is drawn from the Dirichlet distribution with parameter α. The corresponding topic z is generated based on the multinomial distribution with parameter θ. Each word w_n is generated by the distribution $p(w_n|w_n, \beta)$. Finally, we obtain the marginal probability of document w by:

$$p(w|\alpha, \beta) = \int p(\theta|\alpha) \prod_{n=1}^{N} \sum_{z_n=1}^{Z} p(z_n|\theta)p(w_n|z_n|\beta)d\theta \quad (4)$$

where N is size of document and Z is number of topic in document w. In [5], LDA probability of w_i was calculated by combining the topic probabilities with the topic- dependent unigram β:

$$p_{LDA}(w_i) = \sum_{z=1}^{Z} \beta_{w_i,z} \frac{\hat{\gamma}_z}{\sum_{j=1}^{Z} \hat{\gamma}_j} \quad (5)$$

where $\hat{\gamma}_z$ is variational parameter that approximated Bayes estimates for the LDA model via an alternating variational expectation maximization (EM) procedure [7].

2.2 Dirichlet Class Language Model

In [5], Dirichlet class language model (DCLM) is introduced, in which the class structure is estimated by Dirichlet densities from n-gram events. The class uncertainty is compensated by marginalizing the likelihood function over the Dirichlet priors. The latent variable in DCLM reflects the class of an n-gram event rather than the topic in LDA model, which is extracted from large-span documents. DCLM is considered as a kind of class-based LM. In contrast, the class label in a traditional class-based LM has been associated with an individual history word, and derived separately from the stage of model parameters. However, the class structure and the model parameters are consistently estimated under the same criterion in the proposed DCLM.

A linear discriminant function can be used to evaluate the contributions of the historical words to various classes. Without loss of generality, the $(n-1)V$ dimensional history vector h_{i-n+1}^{i-1} is projected into a c dimensional class space using a class-dependent linear discriminant function [8, 9]:

$$g_c(h_{i-n+1}^{i-1}) = a_c^T h_{i-n+1}^{i-1} \quad (6)$$

where a_c is a parameter. This function reflects the class posterior probability $p(c|h_{i-n+1}^{i-1})$, which is essential for predicting the class information for unseen history. $A = [a_1, ..., a_C]$ is a basis vector that in [9] is established to span the class space. DCLM constructs a Bayesian latent class LM by compensating for the uncertainty associated with the latent classes c or class mixtures θ. The class information c in DCLM is drawn from a history dependent Dirichlet prior $\theta = [\theta_1, ..., \theta_C]^T \sim Dir\left(g(h_{i-n+1}^{i-1})\right)$.

The joint probability of word w_i, class c_i and class mixture vector θ conditioned on history h_{i-n+1}^{i-1} and DCLM parameters $\{A, \beta\}$, is computed by:

$$p(w_i, c_i, \theta|h_{i-n+1}^{i-1}, A, \beta) =$$
$$p(w_i|c_i, \beta)p(c_i|\theta)p(\theta|h_{i-n+1}^{i-1}, A) \quad (7)$$

The parameters A, β were estimated using the variational Bayes expectation maximization (VB-EM) algorithm [8]. The n-gram probability obtained using DCLM is expressed in a form of marginal likelihood as:

$$p(w_i|h_{i-n+1}^{i-1}, A, \beta) = \sum_{c=1}^{C} \beta_{w_ic} \frac{g_c(h_{i-n+1}^{i-1})}{\sum_{j=1}^{C} g_j(h_{i-n+1}^{i-1})} \quad (8)$$

Comparing LDA with DCLM indicates that whereas LDA calculates the document probability, DCLM calculates the word probability given history words. DCLM performs the unsupervised learning of latent classes of n-gram events through the VB-EM procedure. DCLM differs from the class-based n-gram, in two ways: firstly, the classes of history words are determined according to the mutual information criterion, secondly, the corresponding classes represent the order of words.

2.3 Cache Dirichlet Class Language Model

In DCLM procedure, the class mixtures θ are drawn from history words w_{i-n+1}^{i-1} using the Dirichlet distribution with parameters $g(h_{i-n+1}^{i-1})$. The class probability $p(c_i|\theta)$ is calculated and the word w_i is predicted incorporating the multinomial parameters $\beta = \{\beta_{w_ic_i}\}$. However, the long-distance information beyond the n-gram window is not captured. In Cache DCLM (CDCLM) [5], in order to perform the large-span

language modeling, the class information must be continuously updated. The class mixtures θ are not only generated from $n-1$ history words but also from the class information $c_1^{i-1} = (c_1, \ldots, c_{i-1})$ of all preceding words w_1^{i-1}. For simplification, CDCLM only used of a best class sequence $\hat{c}_1^{i-1}$. In the new language model, the probability of an n-gram event is calculated as follows:

$$p\big(w_i|h_{i-n+1}^{i-1}, A, \beta, w_1^{i-1}\big) \cong$$
$$\sum_{c=1}^{C} \beta_{ic} \frac{g_c\big(h_{i-n+1}^{i-1}\big)+\rho\sum_{t=1}^{i-1}\tau^{i-t-1}\delta(c,\hat{c}_t)}{\sum_{j=1}^{C}[g_j(h_{i-n+1}^{i-1})+\rho\sum_{t=1}^{i-1}\tau^{i-t-1}\delta(j,\hat{c}_t)]} \quad (9)$$

The derivation of Equation (9) is similar to that of Equation (8). In Equation (9), a weighting factor $0 < \rho < 1$ is empirically introduced to balance the history words and the previous class sequence information. Additionally forgetting factor $0 < \tau < 1$ is applied to discount distant class information. The class associated with the farther word has a smaller impact on the word prediction. In other words, the class sequence is weighted. In the case of $\rho = 0$, CDCLM reduced to DCLM. If ρ is very large, CDCLM is comparable to a class based cache, which is different from the word-based cache in previous cache LMs [10].

3. Proposed Methods

3.1 Dirichlet Class Language Model Based on Part-of-Speech

As mentioned before, the previous adaptation methods extract latent semantic information such as topic dependency of words. In the languages with high morphology for example Farsi, using of the syntactic information such as part-of-speech (POS) along with semantic information can be useful [11].

Accordingly, we proposed an idea for using POS based on DCLM. DCLM acts as a Bayesian topic LM in which the prior density of topic variable is characterized by n-gram events. In our proposed model, we use POS information of previous words along with history words h_{i-n+1}^{i-1} for word clustering. The order of history factors is represented in f_{i-n+1}^{i-1}. Similar to DCLM, first, we declare a linear discriminant function. This function can be used to represent the cooperation of the history factors to different classes.

$$g_c(f_{i-n+1:i-1}^{1:2}) = a_c^T f_{i-n+1:i-1}^{1:2} \quad (10)$$

Linear function shows the class posterior probability $p(c|f_{i-n+1:i-1}^{1:2})$ where $f_i^{1:2}:\{f_i^1 = w_i, f_i^2 = p_i\}$. The first factor is the word and the second one is the POS of the word. a_c is the same as in DCLM.

Fig. 1 shows the graphical model of DCLM based on POS (DCLM_POS) for a text corpus that comprises of previous factor events. The class information c in DCLM_POS is drawn from the parameter θ. The joint probability of word w_i, class c_i and parameter θ, conditioned on history factors $f_{i-n+1:i-1}^{1:2}$ and DCLM_POS parameters {A,β}, is computed by:

$$p(w_i, c_i, \theta|f_{i-n+1:i-1}^{1:2}, A, \beta)$$
$$= p(w_i|c_i, \beta)p(c_i|\theta)p(\theta|f_{i-n+1:i-1}^{1:2}, A) \quad (11)$$
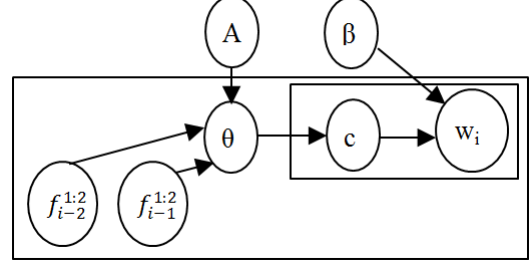
where A, β parameters are the same as in DCLM.

Fig. 1. Graphical representations for DCLM_POS

The n-gram probability based on previous factors is calculated by marginalizing the joint probability over the uncertainty of class mixture θ associated with different classes c_i:

$$p(w_i|f_{i-n+1:i-1}^{1:2}, A, \beta) = \sum_{c=1}^{C} \beta_{w_ic} \frac{g_c(f_{i-n+1:i-1}^{1:2})}{\sum_{j=1}^{C} g_j(f_{i-n+1:i-1}^{1:2})} \quad (12)$$

The DCLM_POS parameters are computed using the VB-EM procedure as in DCLM. After several VB-EM iterations, the DCLM_POS model inference converges.

Comparing LDA in (5) with DCLM_POS in (12) shows that whereas LDA calculates the word probability given word events in documents, DCLM_POS calculates the word probability from history factors.

3.2 Cache Dirichlet Class Language Model Based on Part-of-Speech

From the history factors $f_{i-n+1:i-1}^{1:2}$, the DCLM_POS first draws the class mixtures θ based on the Dirichlet distribution with parameters $g(f_{i-n+1:i-1}^{1:2})$. The word w_i in DCLM is predicted using the multinomial parameters $\beta = \{\beta_{w_ic_i}\}$. DCLM in DCLM_POS was substituted by CDCLM to perform the large-span language modeling and a new CDCLM_POS is developed. In this technique, the class information must be continuously updated. The class mixtures not only depend on $n-1$ history factors but also are influenced by the class information c_{i-2}^{i-1} of all preceding factors $f_{i-2:i-1}^{1:2}$. Fig. 2 shows the graphical model of CDCLM based on POS (CDCLM_POS) using two previous factors for model estimation.
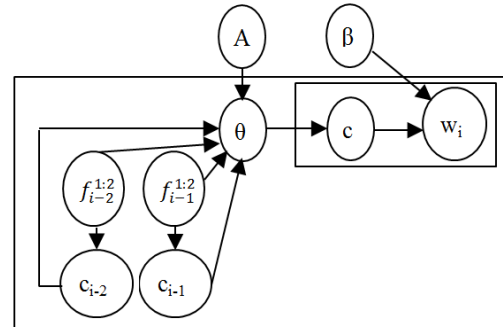

Fig. 2. Graphical representations for CDCLM_POS

As depicted in Fig. 2, to predict w_i, the class mixtures θ, are generated from the history factors $f_{i-2:i-1}^{1:2}$ and the class sequence c_{i-2}^{i-1}. Based on the CDCLM_POS, the probability of an n-gram event is calculated using:

$$p\left(w_i \mid f_{i-n+1:i-1}^{1:2}, A, \beta, w_1^{i-1}\right) \cong$$

$$\sum_{c=1}^{C} \beta_{ic} \frac{g_c(f_{i-n+1:i-1}^{1:2}) + \rho \sum_{t=1}^{i-1} \tau^{i-t-1} \delta(c, \hat{c}_t)}{\sum_{j=1}^{C}[g_j(f_{i-n+1:i-1}^{1:2}) + \rho \sum_{t=1}^{i-1} \tau^{i-t-1} \delta(j, \hat{c}_t)]} \quad (13)$$

where all of parameters are as in Equation (9).

4. Experiments

4.1 Dataset and Experimental Setup

The BijanKhan corpus [12] was utilized to evaluate the proposed methods in continuous speech recognition. The Farsdat training set was adopted to estimate the HMM parameters. The feature vector composed of 12 Mel-Frequency Cepstral Coefficients (MFCC) and one log energy and their first, second and third derivatives. Triphone models were built for 32 phones and each triphone model had three states with sixteen Gaussian mixtures in each state.

The HTK [13] was exploited for HMM training and lattice generation. The baseline LM was trained by SRILM [14] toolkit [1]. Kndiscount [15] method of smoothing methods used in the n-gram model. LDA toolkit[2] to train LDA model and DCLM toolkit[3] to train DCLM based models. In the experiments, number of topics Z and classes C were set to 100. The BijanKhan corpus with 10k documents, 70k distinct words and 40 POS was adopted to train the baseline LM comprised trigram model, FLM and proposed methods [16]. After removing the stop words, we used a lexicon with 45K frequent words for built the LDA model. In addition, the Farsdat corpus is 400 sentences. These corpuses were used to examine different models by perplexity and word error rate (WER). Firstly, we evaluated adapted LMs by perplexity criterion on the 10-fold procedure. Finally, in evaluation of speech recognition, we report WERs (%) of using different LMs.

Perplexity is the most common intrinsic evaluation metric for LM. A lower perplexity corresponds to less confusion in the prediction of language words. We apply 10-Fold mechanism to BijanKhan corpus for perplexity evaluation in all models. The general LM mixture approaches try to combine the topic or class model and traditional LM through some adaptation strategies [17]. Just in the way proposed by [18, 19], many methods are used to integrate the topic or class model with traditional LM which introduces a different type of information.

4.2 Experimental Results

In the experiments, we use linear interpolation for combining the trigram LM and FLM with adaptation methods. The interpolation weight between the basic LM and adaptation methods were determined from the 10-fold mechanism on perplexity metric. Fig. 3 shows the perplexities of the trigram LM, FLM and their linear combination with adaptation methods.
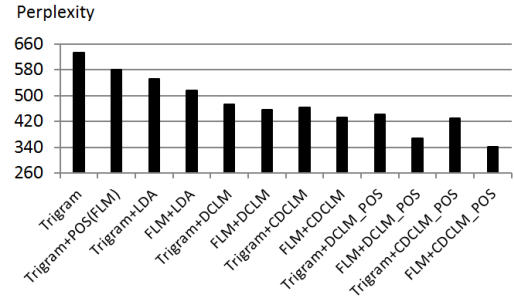

Fig. 3. Perplexities of models with linear interpolation

As Fig. 3 shows, the trigram LM and FLM had perplexity of 632 and 580 respectively. For FLM adaptation models with LDA, DCLM and CDCLM, the perplexities are about 516, 456 and 432 respectively. For FLM adaptation models with the proposed methods, DCLM_POS and CDCLM_POS, the perplexity is reduced to 369 and 342 respectively.

This experiment, represents that language model adaptation with techniques based on DCLM have significant improvement compared with adaptation based on LDA. Furthermore, word clustering has been improved using POS information of history words. In other words, using POS of previous words along with history words and class of history words for word clustering, improves the performance of trigram LM and FLM.

The evaluation of speech recognition was conducted using the Farsdat corpus. We reported the WERs (%) of various LMs. The HTK was used for acoustic model training using HMM and lattice generation. After that, the n-best list is created and then combined with estimated probability produced with LM using linear and log-linear combination. In this experiment, parameter n in n-best list is empirically set to five.

Fig. 4 shows the WERs of linear and log-linear combination of acoustic model with different LMs. This experiment represents that log-linear combination results in less WER than the linear combination.
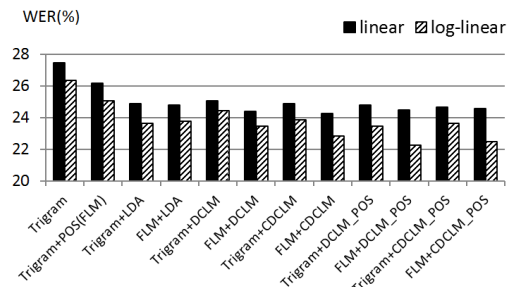

Fig. 4. WERs of combination LMs with acoustic model

---

[1] http://www.speech.sri.com/projects/srilm.
[2] http://www.cs.princeton.edu/~blei/lda-c.
[3] http://chien.csien.ncku.edu.tw/~dclm.

As depicted in Fig. 4, in the log-linear combination, trigram LM and FLM word error rate is about 26.4% and 25.1% respectively. The results of FLM adaptation are better than trigram LM. For FLM adaptation models with LDA, DCLM and CDCLM, word error rate is equal to 23.8%, 23.5% and 22.9% respectively. For FLM adaptation models with the proposed methods, DCLM_POS and CDCLM_POS, WERs is reduced to 22.3% and 22.5% respectively. Therefore, POS information can reduce WER in speech recognition system. Nevertheless, WER results confirm results of perplexity, but the trigram adaptation model with CDCLM_POS has less improvement than DCLM_POS.

The latent classes had been exploited by DCLM_POS methods. These classes were tagged here for ease of understanding. Fig. 5 displays some trigram events samples. Their coordinates in the class space that is spanned by the prior statistics of the previous factors $\{g_c(f_{i-n+1:i-1}^{1:2})\}$ on latent classes consist of POS {"Noun, Noun", "Adjective, Noun"}.

The POS sequences {"Noun, Noun", "Adjective, Noun"} are distributed in the top left and bottom right regions, respectively. The histories that are independent of these two classes are located in the bottom left region. The histories, "حُسن تصادف" and "حَسن تصادف" contained the same word sequences, but were located far apart in the class space.
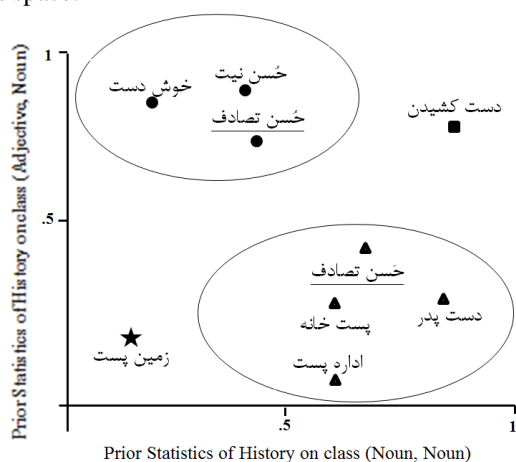
5. Conclusions

In summary, we have compared two approaches for using the part-of-speech (POS) information along with history words and class of history words. We use this information to cluster words and calculate the word probability. The first proposed technique is based on Dirichlet class language model (DCLM) using history words and POS of history words (DCLM_POS). The second is based on cache Dirichlet class language model (CDCLM) using history words, class of history words and POS of previous words (CDCLM_POS). Both methods are combined with trigram language model and factored language model in the form of linear interpolation. In this work, obtained language models are combined with acoustic model for speech recognition. In our experiments, the language model was build using the BijanKhan corpus, and the acoustic model was trained using Farsdat corpus. The lowest perplexity is achieved by linear combination of the factored language model and CDCLM_POS technique. The best word error rate is achieved using log-linear combination of the factored language model and DCLM_POS with acoustic model. As the future work, we will investigate the use of other linguistic features such as morphology. Using these features we hope to get improvements in the large variety of corpus. In addition, we study the discriminant functions in DCLM to reduce its computational complexity for online adaptation.

Fig. 5. Geometrical representations of latent class space constructed by the DCLM_POS

References

[1] L. R. Rabiner; R. W. Schafer, "Theory and application of digital speech processing," Prentice-Hall, 2009.

[2] D. Jurafsky and J. H. Martin, "Speech and Language Processing an Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition," 2nd ed., Prentice Hall, 2008.

[3] K. Kirchhoff, J. Bilmes and K. Duh, "Factored Language Models Tutorial," Department of Electrical Engineering University of Washington, 2008.

[4] D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993-1022, 2003.

[5] J. T. Chien, and C. H. Chuen, "Dirichlet Class Language Models for Speech Recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 3, 2011.

[6] Y. C. Tam, T. Schultz, "Dynamic language model adaptation using variational Bayes inference," *Proc. of EUROSPEECH*, pp. 5-8, 2005.

[7] S. Borman, "The expectation maximization algorithm a short tutorial," Electronic document: www.seanborman.com/publications, 2004.

[8] J. T. Chien, and C. H. Chuen, "Latent Dirichlet Language Model for Speech Recognition," *in Proc. IEEE Workshop Spoken Lang. Technol.*, pp. 201-204, 2008.

[9] J. T. Chien and C. H. Chuen, "Joint Acoustic and Language Modeling for Speech Recognition," *Speech Commun.*, vol. 52, no. 3, pp. 223-235, 2010.

[10] P. R. Clarkson and A. J. Robinson, "Language Model Adaptation Using Mixtures and an Exponentially Decaying Cache," *in Proc. IEEE Int. Conf. Acoustic, Speech, Signal Process.*, pp. 799-802, 1997.

[11] K. Kirchhoff, "Novel speech recognition models for Arabic," JHU, summer workshop final report, 2002.

[12] S. Tasharofi, F. Raja, F. Oroumchian and M. Rahgozar, "Evalution of Statistical Part of Speech Tagging of Persian Text," *International Symposium on Signal Processing and its Applications*, Sharjah, 2007.

[13] S. Young; G. Evermann; D. Kershaw; G. Moore; J. Odell; D. Ollason; D. Povey; V. Valtchev; P. Woodland, "The HTK book," Cambridge University Engineering Department, 2002.

[14] A. Stolcke, "SRILM-An extensible language modeling toolkit," *in Proc. Intl. Conf. Spoken Language Processing*, Denver, Colorado, 2002.

[15] V. Siivola, "Language models for automatic speech recognition: construction and complexity control," dissertations in computer and information science for the degree of doctor of philosophy submitted to the Johns Hopkins University, 2007.

[16] A. Hatami, A. Akbari, B. Nasersharif, "Factored Language Model Adaptation Using Dirichlet Class Language Model for Speech Recognition," *Information and Knowledge Technology (IKT)*, PP. 438-442, 2013.

[17] Z. Lv, W. Liu and Z. Yang, "A New Language Model Adaptation Framework Using Modification of Structures of Background Corpus and Language Model," *Natural Language Processing and Knowledge Engineering*, pp. 1-4, 2009.

[18] j. R. Bellegarda, "Statistical Language Model adaptation: Review and Perspectives," *Speech Communication*, vol. 42, pp. 93-108, 2004.

[19] A. Gutkin, "Log-linear Interpolation of Language Models," Thesis, Department of Engineering, University of Cambridge, UK, 2000.

Ali Hatami received B.S. degree in computer engineering (Software) from the Islamic Azad University, Zanjan, Iran, in 2007. He received the M.S. degree in computer engineering (Artificial Intelligence) from the Iran University of Science and Technology (IUST), Tehran, Iran, in 2012. He is currently Data expert in the Azmoon Keyfiat Company. His research interests include natural language processing, text processing, information retrieval and machine learning.

Ahmad Akbari received the B.S. degree in electronics engineering (1987) and M.S. degree in communication engineering (1989) from the Isfahan University of Technology, Isfahan, Iran. He received the Ph.D. degree in electrical engineering from the University of Rennes, Rennes, France, in 1995. He is Associate Professor in the Department of Computer Engineering, Iran University of Science and Technology. He is currently head of Research Center for Information Technology (RCIT).

Babak Nasersharif received the B.S. degree in hardware engineering from the AmirKabir University of Technology, Tehran, Iran, in 1997. He received M.S. and Ph.D. degree in computer engineering (Artificial Intelligence) from Iran University of Science and Technology, Tehran, Iran, in 2001 and 2007 respectively. He is currently Assistant Professor in the Electrical and Computer Engineering Department K.N. Toosi University of Technology.

PSO-Algorithm-Assisted Multiuser Detection for Multiuser and Inter-symbol Interference Suppression in CDMA Communications

Atefeh Haji Jamali Arani*
Electrical and Computer Engineering Department,Tarbiat Modares University, Tehran, Iran
a.hajijamali@gmail.com
Paeiz Azmi
Electrical and Computer Engineering Department, Tarbiat Modares University, Tehran, Iran
pazmi@modares.ac.ir

Abstract

Applying particle swarm optimization (PSO) algorithm has become a widespread heuristic technique in many fields of engineering. In this paper, we apply PSO algorithm in additive white Gaussian noise (AWGN) and multipath fading channels. In the proposed method, PSO algorithm was applied to solve joint multiuser and inter-symbol interference (ISI) suppression problems in the code-division multiple-access (CDMA) systems over multipath Rayleigh fading channel and consequently, to reduce the computational complexity. At the first stage, to initialize the POS algorithm, conventional detector (CD) was employed. Then, time-varying acceleration coefficients (TVAC) were used in the PSO algorithm. The simulation results indicated that the performance of PSO-based multiuser detection (MUD) with TVAC is promising and it is outperforming the CD.

Keywords: Code Division Multiple Access, Particle Swarm Optimization Algorithm, Multiple Access Interference, Inter-Symbol Interference, Multiuser Detection.

1. Introduction

In direct sequence code-division multiple-access (CDMA) systems, different users transmit signals in the same time and frequency band. Users employ unique spreading codes to be distinguished. Some advantages of utilizing CDMA are: increased capacity, frequency reuse, soft handoff, no need to frequency management or assignment, no guard time, and multipath combating [1]. Direct sequence CDMA (DS-CDMA) is a popular CDMA technique in the wireless communications. The DS-CDMA transmitter multiplies each user's signal by a spreading code. One of the most important problems that limit the capacity of the DS-CDMA systems is the multiple access interference (MAI) that directly relies on the cross-correlations between the spreading codes of all active users. The other problem nominated near-far occurs when the power level of certain users is significantly higher than the others. The near-far problem is an important factor that influences the user capacity and performance of CDMA systems. To obviate this problem, the multiuser detection (MUD) has been introduced as an important method in CDMA systems [2]. In multipath channels, the received signal at the receiver is the combination of delayed versions of the original signal leading to another problem called inter-symbol interference (ISI). The optimum multiuser detector minimizes the probability of error and evaluates a log-likelihood function over the set of all possible users' information sequences. Thus, employing the optimum multiuser detection (OMD) to suppression of MAI (and ISI) in the CDMA systems [3] is an NP-hard problem that increases the computational complexity exponentially with the number of all active users. High computational complexity of OMD leads to considerable efforts in the development of suboptimum detectors with low complexity.

The conventional detector (CD) is one of the suboptimum detectors that are used in CDMA systems. It is to pass the received signal through a bank of filters matched to the users' spreading codes and detects the signal of a user treating the other users' signals as noise. In multipath Rayleigh fading channel, the conventional single user detector is the Rake receiver that combines the outputs of the matched filters that uses the maximum ratio combining (MRC) technique [4].

To solve the NP-hard multiuser detection problems, various optimization algorithms that reduce the complexity of the optimum detector are employed. Particle swarm optimization (PSO) algorithm is one of the optimization algorithms that are applied as an optimization tool in various fields of engineering. Kennedy and Eberhart [5] in 1995 proposed PSO algorithm that is based on the observations of social behavioral models of bird flocking, fish schooling and swarming. It requires the process of initialization to produce the population of random particles. A particle is allocated to a random velocity. Particles fly through the

solution space and a fitness criterion evaluates each of them. Particles accelerate toward those particles that have better fitness values in the swarm. PSO algorithm contains a very simple concept and can be implemented by a few lines of computer code that require only primitive mathematical operators. It is computationally inexpensive in terms of memory and speed as well [6]. At first, PSO was developed for optimization of a continuous variable. Then a discrete binary version of the PSO algorithm was proposed in [7].

In [8], [9] PSO algorithm have been employed to solve the MUD problems in DS-CDMA systems in order to reduce the computational complexity. In [10] binary PSO (BPSO) version was applied to solve the MUD problems. In [11] the de-correlating detector (DD) or linear minimum mean square error (LMMSE) detector was used to initialize the PSO-based MUD. Then, by optimizing an objective function that incorporates the linear system of the DD or LMMSE detector, the PSO algorithm was applied to detect the received data bit in the receiver. However, in these papers ISI problem was not considered. Applying the theory of MUD and evolutionary computation, [12] proposed a hybrid genetic engine, which is suitable for detection of CDMA signals in presence of MAI and ISI. [13] proposed the multiuser detection that uses simulated annealing-genetic algorithm. In [14] a multiuser detector based on tabu simulated annealing genetic algorithm was proposed. In [15] and [16] multiuser detection using quantum clone genetic algorithm and immune clone selection Algorithm were proposed, respectively. Neural network-based CDMA interference cancellation techniques were proposed in [17]. Multiuser detection using hyper differential evolution (H-DE) algorithm was proposed in [18]. In [19] a multiuser detection technique for the DS-CDMA system over a single-path Nakagami-m fading channel is proposed.

In this paper, we considered a DS-CDMA system over additive white Gaussian noise (AWGN) and multipath fading channels. We proposed to apply PSO algorithm to the multiuser detection over multipath Rayleigh fading channel. The conventional detector (CD) is used as the first stage to initialize the PSO algorithm and time-varying acceleration coefficients (TVAC) were used in PSO algorithm. The rest of this paper is structured as follows. In section 2, we introduce our CDMA system model over AWGN and multipath Rayleigh-fading channels, as well as the overview of the CD, DD and optimum multiuser detector. In section 3, the PSO algorithm in AWGN channel and proposed PSO algorithm and its application to CDMA system over multipath channel are given. In section 4, the computation complexity analysis is presented. The simulation results are presented in section 5, and finally conclusions are drawn in section 6.

2.   The Model of DS-CDMA systems

We consider binary phase shift keying (BPSK) transmission through a common AWGN and multipath Rayleigh fading channels shared by K simultaneous users employing a DS-CDMA system. Each user is assigned a normalized spreading code $s_k(t)$ of duration T, where T is the symbol duration. A normalized spreading code can be expressed as:

$$s_k(t) = \sum_{n=0}^{N-1} a_n^k(t) p(t - nT_c) \qquad (1)$$

$$\int_0^T s_k^2(t)\, dt = 1 \qquad (2)$$

Where $[\, a_n^k(t) \in \{+1, -1\}, 0 \le n \le N - 1]$ is the spreading sequence, $p(t)$ is the spreading chip whose duration is $T_c = T/N$. Spreading codes are assumed to be zero outside the interval [0,T]. We describe the model of a CDMA communication system in AWGN and multipath Rayleigh fading channels.

2.1 AWGN channel

We consider a baseband DS-CDMA system over a common AWGN channel, which adds a white random process $n(t)$ to the delayed transmitted signal, with K active users. The baseband received signal $r(t)$ can be written as

$$r(t) = \sum_{k=1}^{K} \sqrt{E_k}\, b_k(t) s_k(t) + n(t) \qquad (3)$$

Where $n(t)$ is white Gaussian noise with power spectral density (PSD) σ^2.

The problem is to observe $r(t)$ and to detect the transmitted bits such that the error probability is minimum. The first step is to reduce $r(t)$ to a set of vector forming a sufficient statistics for b. A sufficient statistic of $r(t)$ is the sampled output of the matched filter (MF) of all the users. Each filter matched to spreading code of a different user. The MF output is then given by

$$y_k = \int_0^T r(t) s_k(t)\, dt \quad , k = 1, \dots, K \qquad (4)$$

Conventional Detector (CD)

CD is to pass the received signal through a bank of filters matched to the users' spreading codes and then to decide on the information bits based on the output

$$\hat{b} = sgn\,[y] = sgn\,[RAb + z] \qquad (5)$$

Where b, z, y denote, the data, noise vector and output of K matched filter, respectively. b, z, y, A and R defined as

$$b = [b_1, b_2, \dots, b_K]^T \qquad (6)$$

$$\begin{cases} z = [z_1, z_2, \dots, z_K]^T \\ z_k = \int_0^T n(t)\, s_k(t)\, dt \quad , k = 1, \dots, K \end{cases} \qquad (7)$$

$$y = [y_1, y_2, \dots, y_K]^T \qquad (8)$$

$$A = diag\,[\sqrt{E_1}, \sqrt{E_2}, \dots, \sqrt{E_K}]^T \qquad (9)$$

$$R = \begin{bmatrix} 1 & R_{1,2} & \cdots & R_{1,K} \\ R_{2,1} & 1 & \cdots & R_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ R_{K,1} & R_{K,2} & \cdots & 1 \end{bmatrix} \qquad (10)$$

Where R is a normalized cross-correlation matrix. The elements of the $K \times K$ matrix R are given by

$$R_{ij} = \int_0^T s_i(t)s_j(t)dt \qquad (11)$$

Where R_{ij} is the cross-correlation coefficient between the ith user's spreading code $s_i(t)$ and jth user's spreading code $s_j(t)$.

De-correlating detector

The CD does not use any information about the other users in CDMA system and therefore can't combat MAI. The DD applies the inverse of the correlation matrix R to the output of the matched filter as shown in Fig. 1.

$$\hat{b} = sgn(R^{-1}y) = sgn(Ab + R^{-1}n) \qquad (12)$$

It is optimum for near-far resistance but it is not optimum in the sense of minimum bit error rate (BER) [**Error! Bookmark not defined.**].
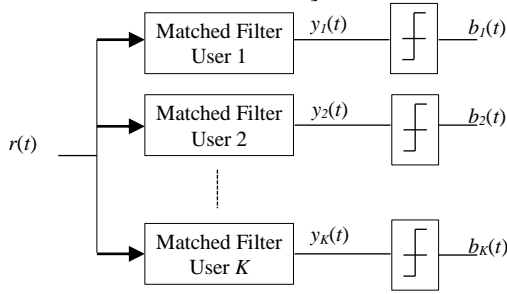

Fig. 1 Decorrelating detector for AWGN channel.

Optimum Multiuser Detection (OMD)

The optimum multiuser detector selects the data sequence b that maximizes the likelihood function that results in the following detection [**Error! Bookmark not defined.**]:

$$b^* = arg\left\{ \max_{b\in\{-1,1\}^K} [2\, y^T Ab - b^T ARAb] \right\} \qquad (13)$$

The OMD rule searches over the 2^K possible combinations of the components b.

2.2 Multipath Fading Channel

We consider BPSK transmission over multipath Rayleigh-fading channels shared by K users employing DS-CDMA, as illustrated in Fig. 2.

The transmitted frame consisting of $2P+1$ number of bits for each user is assumed to be propagating over L independent Rayleigh-fading paths to the base station's receiver.

The impulse response of channel for kth user is

$$h_k(t) = \sum_{l=1}^{L} c_{k,l} e^{-j\phi_{k,l}} \delta(t - t_{k,l}) \qquad (14)$$

Where $c_{k,l}$, $t_{k,l}$ and $\phi_{k,l}$ are the lth path gain, propagation delay and phase for kth user, respectively.

In the Rayleigh-fading channels, the channel gain is a zero-mean complex Gaussian random variable, where the

amplitude $c_{k,l}$ is Rayleigh distributed and the phase $\phi_{k,l}$ is uniformly distributed between $[0,2\pi)$. Both $c_{k,l}$ and $\phi_{k,l}$ are independent for different k and l. The received signal can be written as

$$r(t) = \sum_{i=-P}^{P} \sum_{k=1}^{K} \sqrt{\frac{E_b}{T_b}} b_k(i) s_k(t - iT) * h_k(t) + n(t) \qquad (15)$$

Where the symbol * denotes convolution, $E_k(i)$ and $b_k(i)$ are the power, the ith transmitted information bit of the kth user, respectively; $b_k(i) \in \{+1,-1\}$ is independent equal probability random variable.

$s_k(t)$ is the normalized spreading code of kth user and consists of bipolar rectangular pulses of period with $N\,T_c$, where T_c is the chip period, N is the length of code, such that $N = T/T_c$, and T is the data bit period. AWGN, resulting from receiver thermal noise ($n(t)$), is also considered in this system with PSD σ^2.

The output of the filter matched to the lth path of k at time iT is obtained as

$$y_{k,l}(i) = \int_{-\infty}^{\infty} r(t)s_k(t - iT - t_{k,l})dt \qquad (16)$$

Via matrix notation, the outputs of matched filter can be rewritten as
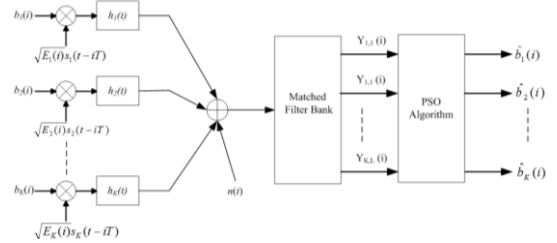
$$Y = RAWB + n \qquad (17)$$

Fig. 2 System model for asynchronous PSO-based multiuser CDMA

Where

$$\begin{cases} Y = (Y(-P), \dots, Y(i), \dots, Y(P)) \\ \quad Y(i) = (Y_1(i), \dots, Y_K(i)) \\ Y_k(i) = (Y_{k,1}(i), \dots, Y_{k,L}(i)) \end{cases} \qquad (18)$$

$$\begin{cases} b = [b(-P), \dots, b(P)]^T \\ b(i) = [b_1(i)\, 1, \dots, b_K(i)\, 1] \end{cases} \qquad (19)$$

Here $1 = (1, \dots, 1)^L$ is an L dimensional vector with all elements equal to 1. A is the matrix of received signal coefficients for all K users defined as

$$\begin{cases} A = diag(\alpha(-P), \dots, \alpha(P)) \\ \alpha(i) = (\alpha_{1,1}(i), \dots, \alpha_{1,L}(i), \dots, \alpha_{K,1}(i), \dots, \alpha_{K,L}(i)) \end{cases} \qquad (20)$$

$$\begin{cases} W = diag(W(-P), \dots, W(P)) \\ \quad W(i) = diag\left(\sqrt{\frac{E_b}{T_b}}\, I_{LK}\right) \end{cases} \qquad (21)$$

Here I_{LK} is an LK-by-LK identity matrix.

Where $\alpha_{k,l} = c_{k,l}e^{-j\phi_{k,l}}$ and R is the correlation matrix,

$$R^{(i)} = \begin{pmatrix} R_{1,1}^{(i)} & \cdots & R_{1,K}^{(i)} \\ \vdots & \ddots & \vdots \\ R_{K,1}^{(i)} & \cdots & R_{K,K}^{(i)} \end{pmatrix} \tag{22}$$

Where

$$R_{k,\acute{k}}^{(i)} = \begin{pmatrix} R_{k,\acute{k}}(1,1,i) & \cdots & R_{k,\acute{k}}(1,L,i) \\ \vdots & \ddots & \vdots \\ R_{k,\acute{k}}(L,1,i) & \cdots & R_{k,\acute{k}}(L,L,i) \end{pmatrix} \tag{23}$$

R represents the cross-correlation functions between the spreading codes of users. $R_{k,\acute{k}}(l,\acute{l},i)$ is $k\acute{k}$th element is obtained as

$$R_{k,\acute{k}}(l,\acute{l},i) = \int_{-\infty}^{\infty} s_k(t - t_{k,l}) \, s_{\acute{k}}(t - t_{\acute{k},\acute{l}} + iT) \tag{24}$$

$R_{k,\acute{k}}(l,\acute{l},i)$ represents the correlation between the kth user's lth multipath component and the $\acute{k}$th user's $\acute{l}$th multipath component.

n is Gaussian noise vector $[n_{k,l}]$ where

$$n_{k,l}(i) = \int_{-\infty}^{\infty} n(t) s_k(t - iT - t_{k,l}) dt \tag{25}$$

The conventional detector is the Rake receiver for each user, which combines the outputs of matched filters using the maximum ratio combining (MRC) method.

$$Y_k = Re\left\{ \sum_{l=1}^{L} \alpha_{k,l}^* \, Y_{k,l} \right\} \tag{26}$$

The optimum multiuser detector selects the data sequence that maximizes the log-likelihood function as follows [20]

$$\Lambda(\mathcal{B}, \alpha) = \sum_{i=-P}^{P} \frac{1}{N_0} (2\Re\{\beta^H B^{(i)} Y^{(i)}\}$$
$$- \sum_{i'=-P}^{P} \beta^T B^{(i)} R^{(i-i')} B^{(i')} \alpha^*) \tag{27}$$

Where

$$\beta_{k,l} = \alpha_{k,l} \sqrt{\frac{E_b}{T_b}} \tag{28}$$
$$\beta = (\beta_{1,1}, \beta_{1,2}, \dots, \beta_{1,L}, \beta_{2,1}, \dots, \beta_{K,L})$$

And $(.)^H$ and $(.)^*$ denote the conjugate transpose and conjugate, respectively.

3. PSO Algorithm

A low-complexity multiuser detector for CDMA systems over AWGN and multipath Rayleigh fading channel by applying the PSO technique is presented in this section. Each solution is represented by a particle in the search space. The particles "fly" or "swarm" through the search space to find the maximum fitness returned by the objective function.

3.1 Parameters of the PSO Algorithm

The parameters of PSO algorithm for the CDMA systems with K users, and AWGN and L multipath channels are briefly stated and defined in the following list.

- **Population size NP:** It is the total number of particles in the PSO algorithm.
- **Particle x_d^{itr}:** It is a candidate solution represented by a K dimensional vector for AWGN channel and $K(2P+1)$ dimensional vector for multipath channel. The dth particle position at the itrth iteration for AWGN channel is defined as

$$x_d^{itr} = [x_{d,1}^{itr}, \dots, x_{d,K}^{itr}] \tag{29}$$

Where $x_{d,k}^{itr}$ is the position of the data bit of kth user of the dth particle.

The dth particle position at the itrth iteration for multipath channel is defined as

$$x_d^{itr} = [x_{d,1,-P}^{itr}, \dots, x_{d,k,i}^{itr}, \dots, x_{d,K,P}^{itr}] \tag{30}$$

Where $x_{d,k,i}^{itr}$ is the position of the ith data bit of kth user of the dth particle.

- **Particle velocity:** since each particle moves it has a velocity v_d^{itr}, where is a K and $K(2P+1)$ dimensional vector for AWGN and multipath channels, respectively. Velocity of particles limited by the $[V^{min}, V^{max}]$.
- **Particle best:** each particle remembers its best position visited so far denoted as $pbest_d^{itr}$.
- **Global best:** each particle knows the best position visited so far among the entire swarm denoted as $gbest^{itr}$.
- **Acceleration coefficients:** c_1 and c_2 represent the weighting of the stochastic acceleration terms that pull each particle toward *pbest* and *gbest* positions. In the original PSO algorithm, the acceleration coefficients are kept constant for all iterations. It was reported in [21] that using time varying acceleration coefficients (TVAC) can enhance the performance of PSO algorithm.

$$c_1 = (c_{1f} - c_{1i}) \frac{itr}{itrmax} + c_{1i} \tag{31}$$

$$c_2 = (c_{2f} - c_{2i}) \frac{itr}{itrmax} + c_{2i} \tag{32}$$

Where c_{1i}, c_{1f}, c_{2i}, and c_{2f} are constants.

- **Objective function:** during each iteration of the algorithm, each solution is evaluated by an objective function to determine its fitness.

We use the log-likelihood function as fitness function. Problem in this paper is to find the particle position to maximize the following cost function Eq.(13) and Eq.(27) for AWGN and multipath channels, respectively.
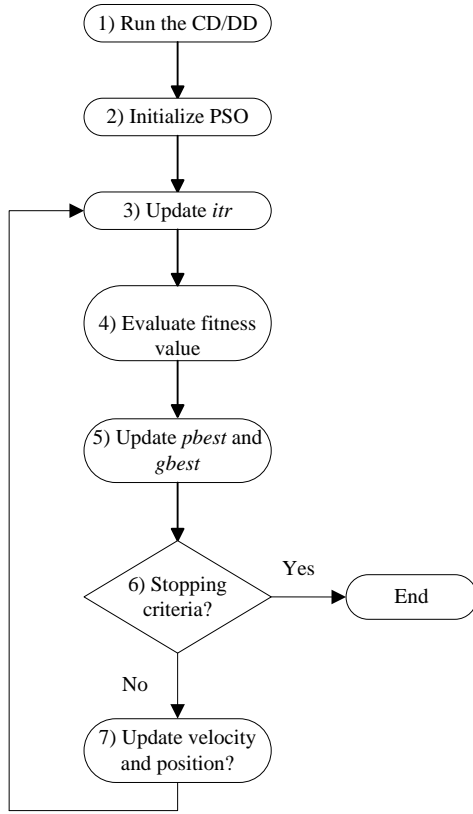
1) Run the CD/DD

2) Initialize PSO

3) Update *itr*

4) Evaluate fitness value

5) Update *pbest* and *gbest*

6) Stopping criteria?

Yes → End

No

7) Update velocity and position?

Fig. 3 Flowchart of PSO algorithm.

3.2 PSO Algorithm

The PSO algorithm flowchart is shown in Fig. 3. The PSO algorithm has the following steps:

1) Run the CD/DD.
2) Initialize the population: a bad initial guess for the PSO algorithm can result in poor performance. In $itr=1$, we set $x_{d=1}^{itr=1}$ to the Rake receiver output using MRC method and conventional/de-correlating detector for multipath fading and AWGN channels, respectively, while the rest of the initial particles $\{x_d^{itr}\}_{d=2:NP}^{itr=1}$ are randomly generated in the search space.
   Evaluate fitness of each particle and update $pbest_d^{itr}$ and $gbest^{itr}$.
3) Update velocity and position: the velocity and position of the *d*th particle in *(itr+1)*th iteration is updated using below equations:

$$v_d^{itr+1} = v_d^{itr} + c_1 * rand(\ ) \\ * \left(pbest_d^{itr} - x_d^{itr}\right) + c_2 \qquad (33) \\ * rand(\ ) * \left(gbest^{itr} - x_d^{itr}\right)$$

$$x_d^{itr+1} = x_d^{itr} + v_d^{itr+1} \qquad (34)$$

In the binary version of PSO algorithm, the particle position is not a real value, but it is either the binary 0 or 1. According to the rule below, we update each component of particle position.

$$if\ rand < S\left(v_d^{itr+1}\right)\ then\ x_d^{itr+1} = 1\ , \\ if\ rand > S\left(v_d^{itr+1}\right)\ then x_d^{itr+1} = -1 \qquad (35)$$

Where $rand(\ )$ and $S(\ )$ are a uniform random number between 0 and 1 and sigmoid function $S(x) = \frac{1}{1+exp(-x)}$, respectively [**Error! Bookmark not defined.**].

4) Stopping criteria: Loop to Step 3) until iterations the maximum number of iteration *itrmax* is reached.

4. Computation Complexity Analysis

It is well known that the computational complexity of the optimum detector in synchronous systems is exponential with the number of active users and in asynchronous systems is exponential with the number of active users and the length of the frame. In synchronous system with *K* active users and the computation Eq.(13) needs Q_G operations containing multiplication and addition operations. The computational complexity of the OMD is $Q_{OMD}=2^K.Q_G$ operations. In asynchronous system, suppose that there are *K* active users and each user transmits $2P+1$ bits and the computation Eq.(27) needs Q_G operations containing multiplication and addition operations. The computational complexity of the OMD is $Q_{OMD}=2^{K(2P+1)}.Q_G$ operations.

With the number of particles *NP* and the maximum number of iterations *itrmax*, the computational complexity of our PSO-based MUD is $Q_{PSO\text{-}basedMUD}=itrmax.NP.Q_G$ operations. Computational complexities of *NP* and *itrmax* have a linear relationship with the dimensionality of the problem of the number of active users *K* and the length of the frame $2P+1$.

The relationship between the computational amounts of two algorithms is

$$Q_{PSO-basd\ MUD} \ll Q_{OMD} \qquad (36)$$

5. Simulation Results

In this section, simulations are presented to evaluate the BER performance results.

5.1 AWGN Channel

We present simulation results to compare the BER performance of the CD, DD and PSO-based MUD. It is assumed that CDMA system supports users' transmission over an AWGN channel with *K=6* and *7* bits gold sequence has been used as the spreading codes.

In Fig. 4, we compare the BER performance of different population size *NP* for $c_1=c_2=2$, and $V_{max}=V_{min}=4$.

As can be observed, the BER of PSO-based MUD is improved with *NP*.

In Fig. 5, The CD and DD output is used as $x_{d=1}^{itr=1}$ (PSO-C and PSO-D in Fig. 5) for initializing of PSO algorithm, respectively. Also, we compare the BER

performance of DD, CD and PSO-based MUD that all of the particles randomly initialized (PSO in Fig. 5).

It is shown that CD and DD have a poor performance. The C-PSO and D-PSO MUD achieve a better performance.
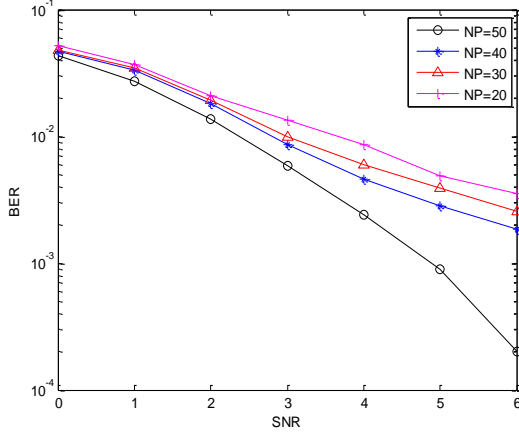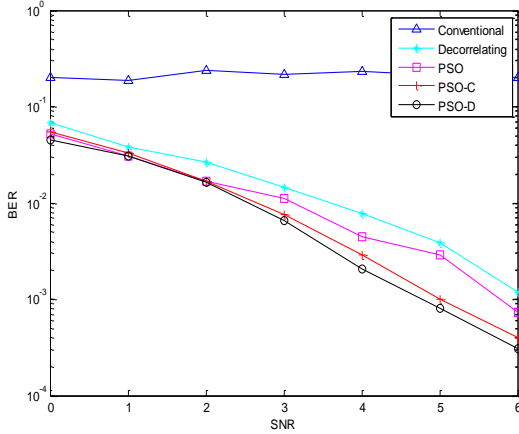

Fig. 4 BER for different population size.

Fig. 5 BER of CD, DD and detection using PSO algorithms.

5.2 Multipath Fading Channel

We consider two users CDMA system to evaluate the BER performance. Assume that the channel has two paths and 7 bits gold sequence has been used as the spreading codes.

1) In the simulation system, $K=2$, $L=2$, $V_{max} = 2, V_{min} = -2, NP = 15, itrmax = 25$.

Under the different SNR, the performance of Rake detector and PSO-based MUD has been shown in the Fig. 6.

As can be observed, the BER of PSO-based MUD is lower than BER of Rake detector.
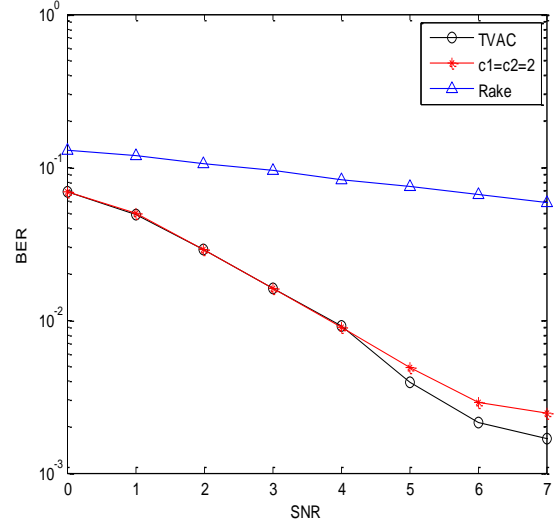
In [22] c_1 and c_2 are selected as follows:

Fig. 6 BER of CD and detection using PSO algorithms.

$$c_1 = (0.5 - 2.5)\frac{itr}{itrmax} + 2.5 \qquad (37)$$

$$c_2 = (2.5 - 0.5)\frac{itr}{itrmax} + 0.5 \qquad (38)$$

Under the different SNR, the performance of PSO-based MUD for $c_1 = c_2 = 2$ and PSO-based MUD for Eq.(37), Eq.(38) has been shown in the Fig. 6.

As can observed, the BER of PSO-based MUD for Eq.(37), Eq.(38) is lower than BER of PSO-based MUD for $c_1 = c_2 = 2$.

2) In Fig. 7, we compare the average BER performance against the number of active users K, given SNR = 8 dB for the asynchronous CDMA system. It can be observed that PSO-based detector has a better BER performance than the Rake detector at the same K.

3) Fig. 8, plots the BER performance of PSO-based and Rake detector against the *Ei/E1* (*i*=2) from 0 to 9 dB, given the number of active user *K*=2 for the asynchronous CDMA system.
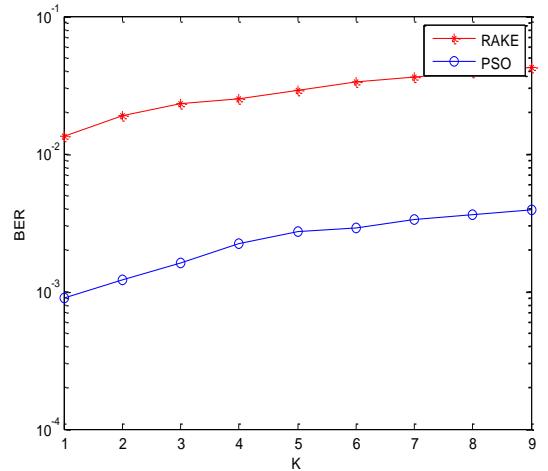

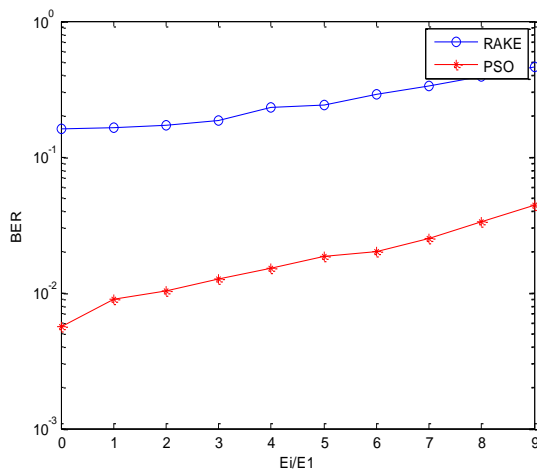Fig. 7 BER against the number of active users *K*, given SNR = 8 dB for the asynchronous CDMA system.

Fig. 8 BER performance against the *Ei/E1* (*i*=2).

6. Conclusions

In this paper, we exploited PSO algorithm in MUD over AWGN and multipath channels. PSO algorithm offered a much lower computational complexity than the OMD algorithm. In the proposed method for joint multiuser and inter-symbol interference suppression in multipath channel, output of the CD was applied as the first stage to initialize the position of a particle. The simulation results showed that the PSO-based MUD has better capability against bit error than CD and using the TVAC can enhance the performance of PSO algorithm.

References

[1] Lee, W. C. Y, "Overview of cellular CDMA". IEEE Trans.Vehi. Tech. Vol . 40, No. 2, 1991, pp. 291−302.

[2] S.Verdu, Multiuser Detection. Cambridge, MA: Cambridge Univ. Press, 1998.

[3] S.Verdu, "Minimum probability of error for asynchronous Gaussian multiple-access channels," IEEE Trans. Inform. Theory, Vol. IT-32, pp. 58-96, Jan. 1986.

[4] J. G. Proakis, Digital Communications. New York: McGraw-Hill, 1989.

[5] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in Proc. IEEE Int. Conf. Neural Networks, Vol. IV, Perth, pp. 1942−1948, Australia, Nov./Dec. 1995.

[6] R. Eberhart and J. Kennedy, "A new optimizer using particles swarm theory," inProc. ISMMHS, 1995, pp. 39–43.

[7] J. Kennedy and R. Eberhart, "A discrete binary version of the particle swarm algorithm," inProc. IEEE CSMC, 1997, pp. 4104–4108.

[8] C. Liue, Y. Xiao, "Multiuser detection using the particle swarm optimization algorithm," proceedings of ISCIT2005, pp. 350-353.

[9] H. L. Hung, Y. F. Huang and J. H. Wen, "A particle swarm optimization based multiuser detector for DS-CDMA communication systems," 2006 IEEE International Conference on systems, Man, and Cybernetics, pp. 1956-1961, Taipei, Taiwan, October 8-11, 2006.

[10] Z. Su Lu and S. Yan," Multiuser Detector Based on Particle Swarm Algorithm," IEEE 6th CAS Symp. on Emerging Technologies: Mobile and Wireless Camm. pp. 783-786, Shanghai, China, May 3 I-June 2,2004.

[11] K. K. Soo, Y. M. Siu, W. S. Chan, L. Yang and R.S. Chen "Particle-swarm-optimization-based multiuser detector for cdma communications," IEEE Trans. Vehicular technology, Vol. 56, No. 5, september 2007.

[12] S. Abedi and R. Tafazolli, "Genetically modified multiuser detection for code division multiple access systems," IEEE Journal On Selected Areas IN Communications, Vol. 20, No. 2, pp. 463-473, FEBRUARY 2002.

[13] Y. Zhou, H. Wang, Y. Wei and J. Wang, "Simulated Annealing-Genetic Algorithm and Its Application in CDMA Multi-user Detection," 3rd International Conference on Intelligent Networks and Intelligent Systems (ICINIS), pp. 638-640, 2010.

[14] Z. Li and D. Ming, "Multi-user detection based on tabu simulated annealing genetic algorithm," International Conference on Instrumentation, Measurement, Computer, Communication and Control (IMCCC), pp. 948-951, 2012.

[15] L. Zhang, L. Zhang and H. Peng, "Quantum clone genetic algorithm based Multi-user detection," The 2nd International Conference on Next Generation Information Technology (ICNIT), pp. 115-119, 2011.

[16] Y. Wang, Z. Feng and L. Jia," A Novel Multiuser Detector Based on Immune Clone Selection Algorithm," Conference on Circuits, Communications and System (PACCS), pp. 238-241, 2010.

[17] B. Geevarghese, J. Thomas, G. Ninan, and A. Francis," CDMA Interference Cancellation Techniques Using Neural Networks in Rayleigh channels," 2013 International Conference on Information Communication and Embedded Systems (ICICES), pp.856-860, Feb. 2013.

[18] Z. Albataineh, F. Salem, " New Blind Multiuser Detection in DS-CDMA Using H-DE and ICA Algorithms," International Conference on 2013 4th Intelligent Systems Modelling & Simulation (ISMS), pp. 569-574, Jan. 2013.

[19] V. K. Pamula, S. R. Vempati, H. Khan, and A. K. Tipparti, " Multiuser Detection for DS-CDMA Systems over Nakagami-m Fading Channels using Particle Swarm Optimization," 2013 IEEE 9th International Colloquium on Signal Processing and its Applications, pp. 275-279, March 2013.

[20] U. Fawer and B. Aazhang, "A multiuser receiver for code division multiple access communications over multipath channels," IEEE TRANSACTIONS ON COMMUNICATIONS, vol. 43, NO. 2/3/4, pp. 1556-1565, FEBRUARY/MARCH/APRIL 1995.

[21] A. Ratnaweera, S. K. Halgamuge and H. C. Watson, "Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients," IEEE Trans on Evolutionary Computation, Vol. 8, No. 3, pp. 240-255. June 2004.

[22] W. Yao, S. Chen, S.Tan and L.Hanzo, "Minimum bit error rate multiuser transmission designs using particle swarm optimisation," IEEE. Trans on Wireless Communications, Vol. 8, No. 10, pp. 5012-5017. October 2009.

Atefeh Haji Jamali Arani received the B.Sc. degree in electrical engineering from Kashan University, Kashan, Iran, and the M.Sc. degree in electrical engineering from Tarbiat Modares University, Tehran, Iran, in 2012. Her research interests are in wireless communications, metaheuristic algorithms and multiuser detection.

Paeiz Azmi was born in Tehran-Iran, on April 17, 1974. He received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Sharif University of Technology (SUT), Tehran-Iran, in 1996, 1998, and 2002, respectively. Since September 2002, he has been with the Electrical and Computer Engineering Department of Tarbiat Modares University, Tehran-Iran, where he became an associate professor on January 2006 and he is a full professor now. Prof. Azmi is a senior mebmer of IEEE.

From 1999 to 2001, Prof. Azmi was with the Advanced Communication Science Research Laboratory, Iran Telecommunication Research Center (ITRC), Tehran, Iran. From 2002 to 2005, he was with the Signal Processing Research Group at ITRC.

Defense against SYN Flooding Attacks: A Scheduling Approach

Shahram Jamali*

Computer Engineering, Associate Professor, University of Mohaghegh Ardabili, Ardabil, Iran

jamali@uma.ac.ir

Gholam Shaker

Young Researchers and Elite Club, Ardabil Branch, Islamic Azad University, Ardabil, Iran

gholamshaker@gmail.com

Abstract

The TCP connection management protocol sets a position for a classic Denial of Service (DoS) attack, called the SYN flooding attack. In this attack attacker sends a large number of TCP SYN segments, without completing the third handshaking step to quickly exhaust connection resources of the victim server. Therefore it keeps TCP from handling legitimate requests. This paper proposes that SYN flooding attack can be viewed metaphorically as result of an unfair scheduling that gives more opportunity to attack requests but prevents legal connections from getting services. In this paper, we present a scheduling algorithm that ejects the half connection with the longest duration, when number of half open connections reaches to the upper bound. The simulation results show that the proposed defense mechanism improves performance of the under attack system in terms of loss probability of requests and share of regular connections from system resources.

Keywords: DoS Attack, SYN Flooding, Scheduling, TCP, Performance

1. Introduction

Security has been always an important issue in communication and computation systems. In these systems security has different aspects. Sometimes the threat is in the form of disclosing of our confident information. In this case we use some security algorithms and protocols to protect the confidentiality. Cryptanalysis [1] and robust key management and distribution algorithms [2, 3] play important roles from this point of view. On the other hand sometimes security is defined as continuous and uninterrupted service. This definition is taken when the threat is in form of a DoS attack. The goal of a DoS attack is to completely tie up certain resources so that legitimate users are not able to access a server. A successful DoS attack overpowers the victim and conceals the offender's identity [4]. A DoS attack can be regarded as an explicit attempt of attackers to prevent legitimate users from gaining a normal network service. DoS attacks typically trust on the misuse of exact susceptibility in such a way that it consequences in a denial of the service. New arithmetical assessment show that DoS positions at the quarter place in the list of the most poisonous attack classes in contradiction of information systems [5]. Anderson et al. rely on the use of a 'send-permission-token' to restrict DoS attacks [6]. Allen and Marin use estimates of the Hurst parameter to identify attacks which cause a decrease in the traffic's self-similarity [7]. Two statistical methods of analyzing network traffic to find DOS attacks are provided in [8]. One monitors the entropy of the source addresses found in packet headers, while the other monitors the average traffic rates of the 'most' active addresses. Many proposals have been made for IDSs intended to detect DoS attacks [9-12], most of them being based on the arithmetical detection of high traffic rate spending from the interloper or interlopers.

Generally DoS attacks could have two major forms. In the first one, the malicious user crafts very carefully a packet trying to exploit vulnerabilities in the implemented software (service or a protocol). In the second form, the malicious user is trying to overwhelm system's resources of the provided service-like memory, CPU or bandwidth, by creating numerous of useless well-formed requests. This type of attack is well known as flooding attack [13]. One of the most common DoS flooding attacks is SYN flooding attacks. It works at the conveyance layer. A TCP connection is recognized in what is known as a 3-way handshake. When a client labors to start a TCP connection to a server, first, the client needs a connection by distribution an SYN packet to the server. Then, the server returns a SYN–ACK, to the client. Lastly, the client admits the SYN–ACK with an ACK, at which point the connection is recognized and data transfer commences. In an SYN flooding attack, attackers use this protocol to their advantage. The attacker directs a large number of SYN packets to the server. All of these packets have to be touched like a connection request by the server, so the server must response with a SYN–ACK. The attacker then has two choices. One is just not to response to the SYN–ACK, which will reason the server to have a half-open connection. This would let the server to chunk any further packets from the attacker's IP address, ending the attack hastily. Then again, the attacker parodies the IP address of some unwary client. The server rationally responses to this IP address, but the genuine client really

---

* Corresponding Author

exist in at this IP address will weakening this SYN–ACK as it did not pledge the connection. The result is that the server is left waiting for a reply from a large amount of connections. Since reserve of any system is imperfect, then, there are a limited number of connections a server can handle. Once all of these are in use, waiting for connections that will not ever come, no new connections can be made whether valid or not. It is clear that though this is a conveyance layer attack, it touches all TCP-based requests in the victim server. When the server cannot handle new connections, any request that tries to establish TCP connections with the server, fails in its effort. Note that SYN flooding attacks goal to use TCP buffer space and do not touch the parameters such as link bandwidth, dispensation capitals and so on.

We believe that in a SYN flooding attack an unfair scheduler gives more opportunity to attack requests but prevents legal connections from getting service. Hence, we propose a scheduling-based solution in which when an arriving request faces with a full queue, the oldest half connection is terminated and its resources are assigned to the new connection.

2. Related Work

There are many researches focusing on SYN flooding attack. S.H.C. Haris et al. in [14] used anomaly detection to detect TCP SYN flooding attack based on payload and unusable area. In order to detect TCP SYN flooding, the normal payload characters must be understand first unless the analysis will take times for those that are not expert in payload characters. Other researches [15, 16] proposed methods to throttle spoofed SYN flooding attacks at the sources. Long et al. [17] proposed two queuing models for the DoS attacks in instruction to get the pack postponement jitter and the loss probability. Wang et al. [12] educations the DoS attacks logically by using a more general queue model, a two-dimensional embedded Markov chain, which can more precisely capture the dynamics of the actual DoS attacks. Maciej Korczynski et al. [18] proposed an accurate sampling scheme for defeating SYN flooding attacks as well as TCP port scan activity. The scheme examines TCP segments to find at least one of multiple ACK segments coming from the server to validate legitimate connections. The method achieves good detection performance with false positive rate close to zero even for very low sampling rates. Hussain and Blazek assessed traffic arrivals and consistent ramp-up doings in [12, 19].Their approaches try to learn how a signal in a system and parameters explaining the system change. To detect attacks, packet rate against time is examined instead of only the packet header. This is done, in instruction that IP address spoofing cannot deceive the attack detection method. Haidar Safa et al. in [20] proposed a novel defense mechanism that makes use of the edge routers that are associated with the spoofed IP addresses' networks to

determine whether the incoming SYN–ACK segment is valid. This is accomplished by maintaining a table that matches the incoming SYN–ACKs with outgoing SYNs and also by using the ARP protocol. Other investigates made in this area can be deliberate in [21, 22]. As another research S. Jamali and his coworkers have presented a PSO-based defense scheme against SYN flooding attacks [23], called PSO-SFDD. This algorithm formulates the defense issue as an optimization problem and then employs PSO algorithm to solve it. The achieved solution is a configuration for the under attack system that alleviates the attack effects.

3. Scheduling Algorithms

When a computer or network resource receives multiple service requests (jobs) at a given time, a scheduling algorithm is necessary to determine the order in which requests are serviced. Scheduling algorithms work based on several job features such as processing time, priority, due date, and so regarding their design goals [24]. In this section, we first present some of these scheduling algorithms and then present a novel scheduling algorithm which is used to design a defense strategy against SYN flooding attacks.

3.1 A Review over Scheduling Algorithms

There are many different scheduling algorithms proposed for scheduling in different operating systems or switches. Below we cover some popular algorithms, although there are some other algorithms such as Round Robin, Priority Scheduling, Multi-level Queue Scheduling and Real Time Scheduling [24] which will not be covered here.

FIFO-A common method of job scheduling for computer and network resources is First-Come-First-Serve (FCFS) or FIFO, where jobs are serviced in the order in which they arrive. Using FIFO, the job with the earliest arrival time is served first. Jobs with earlier arrival times are served before jobs that arrive later [25].

SPT (Shortest Processing Time)-Using SPT, jobs are processed in ascending order of processing times. It is well known that SPT minimizes the total completion times of a set of jobs. SPT produces an optimal job sequence for minimizing the total, and thus mean, of job waiting times [25, 26].

Hard fairness-Hard fairness [27] is also known as round-robin scheduling. Homogeneity of resources is not a requirement in this type of scheduling. It is the fairest scheme since each process is guaranteed exactly equal amount of time in order.

Max–Min fairness-Max–min fairness [28] allocates resources in order of increasing demand. The minimum amount of resources assigned to each process is maximized. So if there are more than enough resources

for each process, every process gets what it needs. If there is not, the resources are split evenly. This means that the process which require fewer resources get a higher proportion of their need satisfied. This type of scheme works best when there is not large differences in amount of resources requested by different processes.

SRTF (Shortest Remaining Time First)-Shortest Remaining Time First, also known as **Shortest Job First (SJF)**, is a scheduling method that is a preemptive version of shortest job next scheduling. In this scheduling algorithm, the process that needs the smallest amount of remaining time to be completed is selected to execute. Since the currently executing process has the shortest amount of remaining time and since that time should only reduce as execution progresses, processes will always run until they complete or a new process is added that requires a smaller amount of time.

3.2 Proposed Scheduling Algorithms: Highest Residence Time Ejection (HRTE)

This algorithm is a preemptive two-phase scheduling algorithm. This algorithm is useful for scenarios in which the service time of requests is unknown. According to this scheduling algorithm, while input queue isn't full, HRTE is in its first phase and acts exactly like round robin algorithm. But upon queue becomes full and arriving requests are blocked, HRTE switches to its second phase during which ejects the job with the highest residence time and assigns the released capacity to the arriving requests. HRTE remains in this phase until a free capacity is available in the wafting queue. In other words, it has some similarities with SRTF, but when remaining time is unknown SRTF cannot be used. In this situation we assume that those requests that have the longest duration in the past will have the longest remaining time, as well and hence will be rejected.

4. Proposed Defense Scheme: Queue Scheduling

When a connection request arrives at a TCP-based server, receives a buffer space of the backlog queue upon finding an inactive buffer space and is blocked otherwise. Now, consider a server under the SYN flooding attacks. Assume that in this computer each half-open connection is held for at most a period of holding time (h), and at most a number of maximum half-open connections (m) are allowed. We assume that a half-open connection for a regular request packet is held for a chance time which is exponentially distributed with parameter μ. The arrivals of the regular request packets and the attack packets are both Poisson processes with rates λ1 and λ2, respectively. The two arrival processes are independent. Obviously, when the system is under attack then number of pending connections increases and in a point in which there is no more room for pending connection to be saved the arriving requests will be blocked. In the other words, when a server is under SYN flooding attacks, half-open

connections quickly consume all the memory allocated for the pending connections and prevent the victim from further accepting new requests, leading to the well-known buffer overflow problem.

We believe that the defense against this attack can be considered as a queue scheduling algorithm that differentiates attack requests from regular requests and then ejects the attack requests. This defense scheme, called HRT_SYN, tries to block attack connections and to prevent the system from allocating buffer space to attack connections. To this end, we use the proposed scheduling algorithm and keeps attack request from occupying system resources for a long time. According to this approach while there is free capacity for coming connection requests they will be accepted and inserted in the queue. But, when a connection arrives and faces with a full queue, then the HRTE scheduling algorithm is invoked and ejects the connection with the highest residence time. The ejected connection is likely an attack request that leaves the system to make the capacity free. This defense scheme can be described by flowchart of Figure 1. To measure efficiency of defense algorithms we use three parameters i.e. RT, AT and Ploss:

RT is Regular requests residence Time that can be described as mean ratio of the occupied resources by regular connections to total available connection resources. It is computed by using equation (1).

$$RT = \frac{\sum_{i:\ \text{all regular connections}} \text{duration}_i}{\text{simulation time} * m} \qquad (1)$$

In which, *m* is maximum number of connections that are allowed to be established in the system.

AT is Attack residence Time that can be described as mean ratio of connection opportunistic that is occupied by Attack connections. It is computed by using equation (2).

$$RT = \frac{\sum_{i:\ \text{all attack connections}} \text{duration}_i}{\text{simulation time} * m} \qquad (2)$$

Ploss *(Probability of connection loss)* is connection loss probability, a basic measure for assessing the performance of the system under DoS attacks. Each arriving connection request packet is rejected once there are *m* pending connections in the system. On the other hand some half-open connections are ejected after a period of time (in Linux and PSO_SFDD) or when there are *m* pending connections and a new connection request arrives (in HRT_SYN). Therefore *Ploss* can be described as ratio of the number of ejected or rejected requests to the all arrived requests.
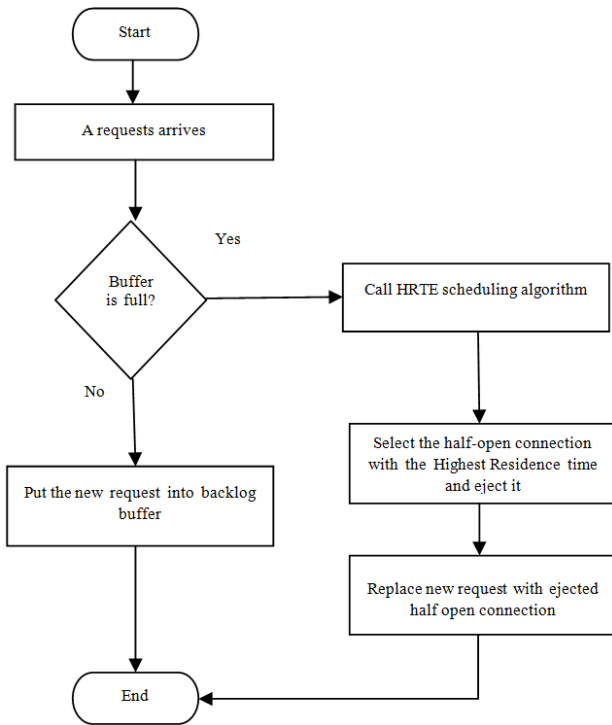
Fig. 1. HRT_SYN in operation

5. Implementation Issues and Simulation Results

In his section we use MATLAB software package to gather a group of simulation results to demonstrate the validity of our design. In this study TCP of the Linux operating system along with PSO-SFDD are considered as a comparison reference point to show efficiency of HRT_SYN algorithm.

5.1 Simulation Setup

As an important issue, note that the victim server doesn't have any information to determine whether the arriving requests are regular requests or attack requests. It accepts all arriving requests and then applies scheduling algorithm. Arrival rate of regular requests to the server has Poisson distribution with mean rate λ and attack requests has mean rate $k\lambda$, in which, k represents the ratio between arrival rate of the attack requests to arrival rate of the regular requests and is referred to as attack intensity. Duration of half-open connections for normal requests has exponential distribution with mean μ. In this simulation we consider a server in a network with $\lambda=10/s$ and $\mu=100$ s. Attack intensity i.e. k is set with different values in different simulation scenarios.

5.2 Simulation Results

In this section a typical TCP protocol is equipped with HRT_SYN defense capabilities and then is compared with Linux TCP which uses static values for maximum

holding time and maximum number of half open connections (75 second and 125 connections respectively).

In this study three scenarios of simulations are presented. In the first scenario attack intensity is considered as k=0.1, in the second one it is considered as k=3 and finally in the third scenario, we consider a variable attack intensity that fluctuates between 0 and 3 as shown in Figure 2.
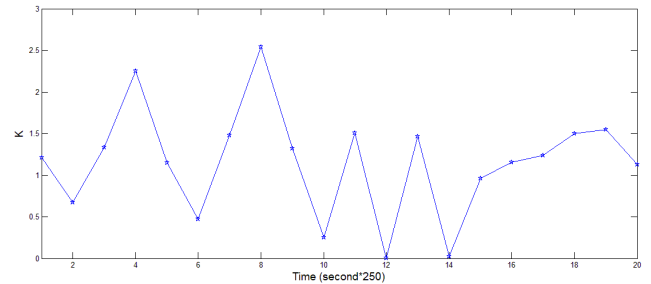

Fig. 2. Variable intensity of the attack in scenario 3
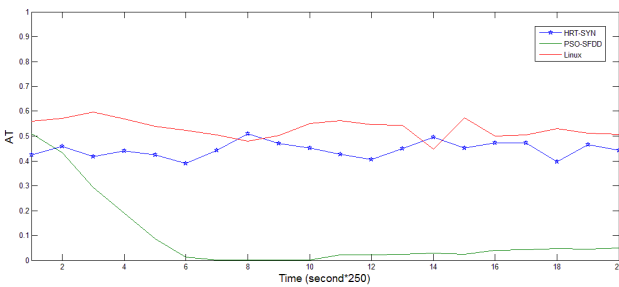
Scenario 1: Low attack intensity (k=0.1)

In order to study behavior of HRT _SYN against a low rate attack, we consider a scenario in which attack intensity is considered as k=0.1. Simulation results are given in Figure 3. This figure shows that defense performance of HRT_SYN are as performance of Linux TCP. According to Figures 3.a Ploss is zero for HRT_SYN, PSO_SFDD and Linux TCP during the simulation time. This means that since attack requests have low rate in this scenario, backlog buffer will never be full and hence no request will be blocked. Figures 3.b and 3.c shows interesting differences among HRT_SYN and Linux TCP in one side and PSO_SFDD on the other side. PSO_SFDD decreases both AT and RT. This has root in this fact that when PSO_SFDD faces with any attack connections it strictly reduces the allowed residence time of all half open connections. Obviously this affects both regular and attack connections and hence reduces both RT and AT. But since in this scenario attack rate is low always there is some available unused capacity for coming connection requests, and HRT_SYN will not be activated to eject the oldest half-open connection and hence RTT and AT will be increased.
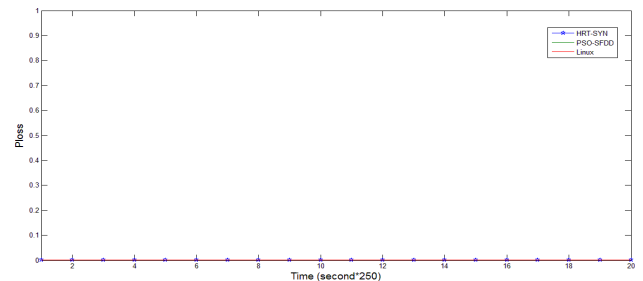
a. Connection Requests Loss Probability
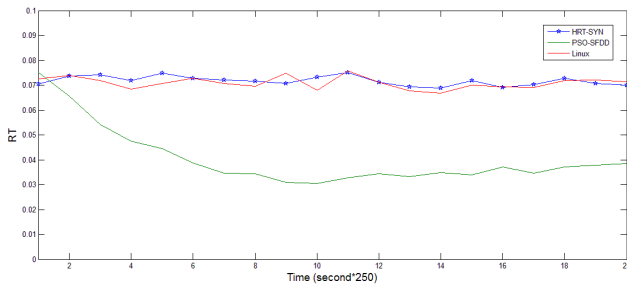
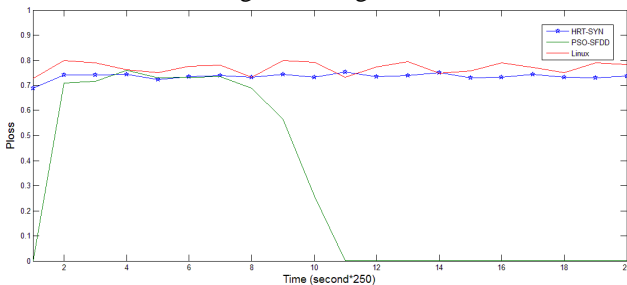b. Regular Requests Residence Time
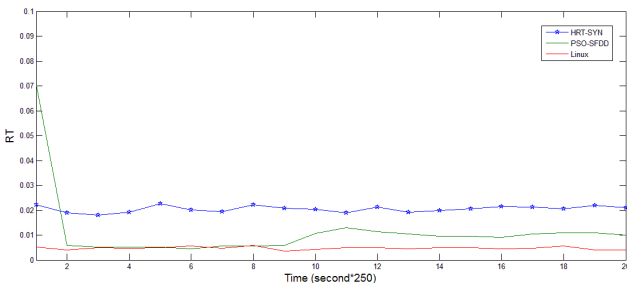

c. Attack Requests Residence Time

Fig. 3. Comparision of PSO_SFDD, HRT_SYN and Linux TCP
for k=0.1

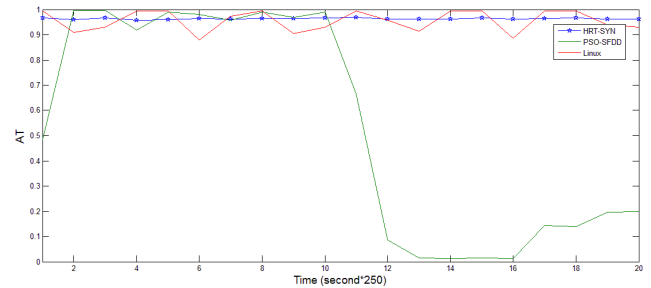**Scenario 2: High attack intensity (k=3)**

In this scenario the victim server goes under a heavy SYN flooding attack in which the number of attack requests is thrice the number of regular requests. Simulation results are given in Figure 4.

a. Connection Requests Loss Probability

b. Regular Requests Residence Time

c. Attack Requests Residence Time

Fig. 4 Comparision of PSO_SFDD, HRT_SYN and Linux TCP
for k=3

This figure shows that PSO_SFDD outperforms Linux TCP and HRT_SYN in terms of Ploss and AT, but the best RT is achieved by HRT_SYN. Since HRT_SYN eject only the connection with highest residence time, it likely affects only attack connections and hence it increases share of regular connections share from connection capacity. This is due to the fact that HRT_SYN terminates those half-open connections that have been established by attackers and occupy the victim resources for lomg. As shown, HRT_SYN has lower *Ploss* in compare with Linux TCP, because it doesn't allow attack requests to remain inside the system and ejects them to provide free capacity for coming connections requests. Since PSO_SFDD decreases maximum holding time of half open connections, attack connection will be terminated sooner and hence incoming requests will have more chance to establish a connection. This is the reason for lower *Ploss* of PSO-SFDD in Figure 4.

Scenario 3: Variable attack intensity (k fluctuates between 0 and 3)

In this scenario, the attack intensity changes during the time. As shown in Figure 2 the attack intensity oscillates in range of [0, 3] during the simulation time. Figure 5 shows that HRT_SYN improves the server performance in term of RT in compare with PSO_SFDD and TCP Linux. But as the previous scenario PSO_SFDD outperforms Linux TCP and HRT_SYN in terms of *Ploss* and AT. Note that RT shows share of regular connections from TCP resources and hence, it can be said that HRT_SYN provide better performance and serves regular connections more efficiently.

6. Conclusions

This paper proposed a scheduling-based approach to defend against SYN-flooding DoS attacks. We used a simple queuing model, to show important metrics of a computer system which is under SYN flooding attacks. Then, we used scheduling algorithm to draw a successful defense against SYN flooding attack requests. This scheduling algorithm ejects the half connection with the longest duration, when number of half open connections reaches to the maximum value. Simulation result showed that the proposed defense strategy behaves extremely

better than Linux TCP and PSO-SFDD algorithms. Since HRT_SYN accepts all arrived connection requests and ejects the oldest half open connection its success is reflected by RT.

As a direction for future works we can focus to design a more efficient scheduling algorithm to provide better performance for the scheduling based defense scheme. This scheduling algorithm can consider other parameters such as statistical characteristics of previous regular and attack connections.
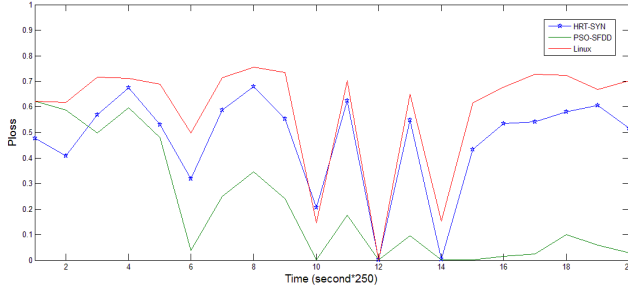
b. Regular Requests Residence Time

a.Connection Requests Loss Probability
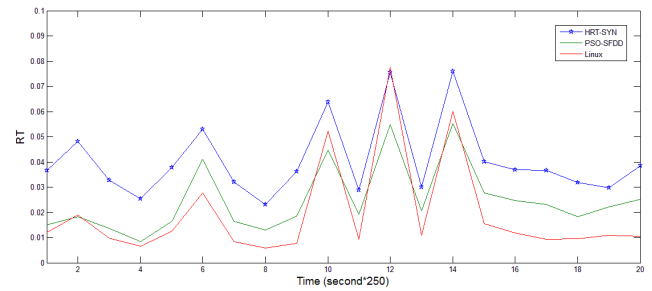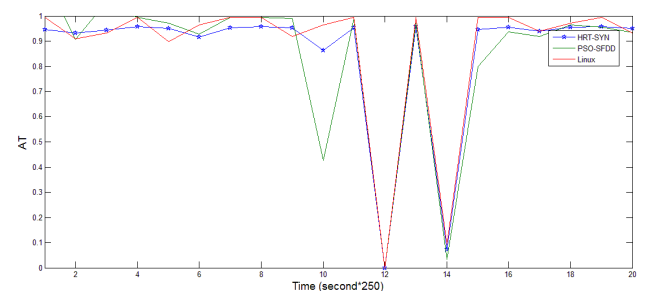
c. Attack Requests Residence Time

Fig. 5. Comparision of TCP Linux, PSO_SFDD and Linux TCP for variable traffic intensity k= [0, 3]

References

[1] X. Wang, W. Guo, W. Zhang, M. Khurram Khan, Cryptanalysis and improvement on a parallel keyed hash function based on chaotic neural network, Telecommunication Systems, 2013.

[2] J. Kim, K. Kim, A scalable and robust hierarchical key establishment for mission-critical applications over sensor networks, Telecommunication Systems, 2013.

[3] J. Teo, Ch. How Tan, J. Mee Ng, Denial-of-service attack resilience dynamic group key agreement for heterogeneous networks, Telecommunication Systems, 2007.

[4] Sh. Chen, Y. Ling, R. Chow, Y. Xia, AID: A global anti-DoS service, Computer Networks, Vol. 51, 2007.

[5] L.A. Gordon, M.P. Loeb, W. Lucyshyn, R. Richardson, CSI/FBI computer crime and security survey, Computer Security Institute, 2005.

[6] T. Anderson, T. Roscoe, D. Wetherall, Preventing internet denial-ofservice with capabilities. Computer Communications Review, 2004.

[7] W. Allen, G. Marin, The LoSS technique for detecting new denial of service attacks. In: SoutheastCon; 2004.

[8] L. Feinstein, Statistical approaches to DDoS attack detection and response. In: DARPA information survivability conference and exposition proceedings, 2003.

[9] R. B. Blazek, H. Kim, B. Rozovskji, A. Tartakovsky, A novel approach to detection of denial of service attacks via adaptive sequential and batch sequential change point detection methods, in: IEEE Workshop on Information Assurance and Security, 2001.

[10] P. Mell, Donald Marks, Mark McLarnon, A denial-of-service resistant intrusion detection architecture, Computer Networks, 2000.

[11] A. Nadeem, M. Howarth, Protection of MANETs from a range of attacks using an intrusion detection and prevention system, Telecommunication Systems, 2013.

[12] Y. Wang, Ch. Lin, Q.L. Li, Y. Fang, A queuing analysis for the denial of service (DoS) attacks in computer networks, Computer Networks, 2007.

[13] D. Geneiatakis, N. Vrakas, C. Lambrinoudakis, Utilizing bloom filters for detecting flooding attacks against SIP based services, Computers & Security, 2009.

[14] S.H.C. Haris, R.B. Ahmad, "M.A.H.A. Ghani,"Detecting TCP SYN Flood Attack based onAnomaly Detection," Second International Conference on Network Applications, Protocols and Services, 2010.

[15] W. Chen, D. Yeung, Throttling spoofed SYN flooding traffic at the source, Telecommunication Systems, 2006.

[16] B. Xiao, W. Chen, Y. He, An autonomous defense against SYN flooding attacks: detect and throttle attacks at the victim side independently, Parallel and Distributed Computing, 2008.

[17] M. Long, C.H. Wu, J.Y. Hung, Denial of service attacks on network-based control systems: impact and mitigation, IEEE Transactions on Industrial Informatics, 2005. [6] P. Mel, D. Marks, M. McLarnon, A denial-of-service resistant intrusion detection architecture, Computer Networks , 2000.

[18] M. Korczy´nski, L. Janowski, and A. Duda, "An Accurate Sampling Scheme for Detecting SYN Flooding Attacks and Port scans," IEEE ICC, 2011.

[19] A. Hussain, J. Heidemann, C. Papadopoulos, A framework for classifying denial of service attacks, USC Information Sciences Institute, 2003.

[20] H. Safa, M. Chouman, H. Artail, M. Karam, A collaborative defense mechanism against SYN flooding attacks in IP networks, Network and Computer Applications, 2008.

[21] M. Hamdi, N. Boudriga, Detecting denial-of-service attacks using the wavelet transform, Computer Communications, 2007.

[22] V.A. Siris, F. Papagalou, Application of anomaly detection algorithms for detecting SYN flooding attacks, Computer Communications, 2006.

[23] S. Jamali, G. Shaker, PSO-SFDD: Defense against SYN flooding DoS attacks by employing PSO algorithm, Computers and Mathematics with Applications, 2012.

[24] P. Brucker, Scheduling Algorithms, Fifth Edition, Springer-Verlag, 2006.

[25] M. Pinedo, Scheduling theory, algorithms and systems. Englewood Cliffs, NJ: Prentice-Hall; 1995.

[26] S. Eilon, I. Chowdhury, "Minimizing waiting time variance in the single machine problem", Management Science, 1977.

[27] D. Koutsonikolas, S.M. Das, Y.C. Hu, An interference-aware fair scheduling for multicast in wireless mesh networks, Journal of Parallel and Distributed Computing, 2008.

[28] S. Singh, U. Madhow, E.M. Belding, Beyond proportional fairness: a resource biasing framework for shaping throughput profiles in multi hop wireless networks, INFOCOM, 2008.

Shahram Jamali is an associate professor leading the Autonomic Networking Group at the department of computer engineering, University of Mohaghegh Ardabili. He teaches on computer networks, network security, computer architecture and computer systems performance evaluation. Dr. Jamali received his M.Sc. and Ph.D. degree from the Dept. of Computer Engineering, Iran University of Science and Technology in 2001 and 2007, respectively. Since 2008, he is with Department of Computer Engineering, University of Mohaghegh Ardabil and has published more than 60 conference and journal papers.

Gholam Shaker received his B.Sc. from Payam Noor University Ardabil, Iran, a M.Sc. (2011) from IAU (Islamic Azad University-Zanjan branch) University. All are in computer engineering. He is currently teaching in department of computer science at IAU University. His research interests include Quality of service, security systems and communication networks.