International Journal of
Intelligent Systems and
Applications in Engineering
ISSN:2147-6799 www.atscience.org/IJISAE
Advanced Technology and Science
Original Research Paper

Fuzzy Multicriterial Methods for the Selection of IT-Professionals

Masuma Mammadova[1,*], Zarifa Jabrayilova[1], Faig Mammadzada[1]

Received 05th September 2014, Accepted 18th December 2014

Abstract: This paper presents the solution of issues related to selection based on evaluation of demand set forth to IT specialists, to develop appropriate decision support system. In this case problem is reduced to multicriterial task of decision making, functioning in a fuzzy environment. We propose criteria estimation method allowing regulation and selection of the best alternative according to the scenario appropriate to the requirements of the decision making person, at a current time. For realization of abovementioned task on the basis of fuzzy logic methods we propose methods of expert knowledge processing of the importance criteria and their characterizing factors.

Keywords: *Decision Support System, Fuzzy Environment, Personnel Selection Problem, Fuzzy Multicriterial Model, Multiscenario Approach, Importance Factor of the Criteria*.

1. Introduction

Dynamics of expansion of ICT sector, state policy and strategic documents in this area, oriented to integration of the country into global information space, Azerbaijan IT-industry growth rates have caused sharply increased demand for IT-professionals. According to the results of monitoring of a supply and demand in the labour market of the IT-professionals, carried out by Institute of information technologies, for today the ratio of demand for IT-professionals exceeds corresponding offers approximately in 3 times, i.e. IT-professionals in the market are required three times more, than train the high schools [1]. Moreover the diversification of Azerbaijan economy stipulates the further expansion and a deepening of IT applicable spheres and enables to assume, that there is no fear of the fall of demand for IT-professionals nearest years.

Penetration IT into the diversified spheres of human activity promotes diversification of the IT-segment, that, on the one hand, causes the transformation of old IT-professions, on the other hand, stimulates occurrence in the market the new ones.

Within the framework of "Monitoring of supply and demand in IT labour market of Azerbaijan", the list of IT professions and their ranking based on evaluation of demand of IT professions was determined [1], [2].

Current article reviews the solution of issues related to selection based on evaluation of demand set forth to IT specialists. Reviewed problem for this purpose was resolved within multicriteria selection problem. References [3] and [4] reviewed the personnel selection studies and found that the several main factors including change in organizations, change in work, change in personnel, change in the society, change of laws, and change in marketing have influenced personnel selection. In literature, there are a number of studies which use heuristic methods for employee selection.

A fuzzy MCDM framework based on the concepts of ideal and anti-ideal solutions for the most appropriate candidate is presented in [5]. Also, a fuzzy number ranking method by metric distance for personnel selection problem was proposed in [6] and a personnel selection system based on fuzzy AHP was developed in [7].

In addition, researchers used fuzzy technique for order preference by similarity (TOPSIS) based on the veto threshold for ranking job applicants [8]–[10].

Recently, owing to the advancements in information technology, researchers have developed decision support systems and expert systems to improve the outcomes of HRM [11], [12].

A model to design an expert system for effective selection and appointment of the job applicants developed in [13]. Although the applications of expert system or decision support systems on personnel selection and recruitment are increasing [14]–[16], however, the research taking into account requirements of the employer in the real time has not been considered in those papers.

2. Characteristic aspects and conceptual model of recruitment issues related to IT-professions

The list of criteria for recruitment as an IT professional, set forth by the employers for those wishing to be employed have been determined. Criteria are presented in 6 groups: criteria are presented as following, K_1 – age, K_2 – gender, K_3 – education, K_4 – personal qualities, K_5 – professional requirements in IT specialization, K_6 – additional capabilities. Each of these criteria is defined by multiple indicators that characterize them [17].

One of the complication problems during the solution of this issue, is determination of knowledge and capabilities of the job applicant in accordance with professional requirements and determination of his/her suitability level to requirements set forth to occupy this position. I.e. above listed are determined through multiple indicators with different importance levels. For instance, it is necessary to determine the level of personal qualities of the job applicant for IT position, such as performance discipline, initiative at work, capability to pass on experience, team work (communication) capability and analytical thinking, and find their importance coefficient with regard to each other; which requires attraction of experts to the process.

[1] *Institute of Information Technology of National Academy Science of Azerbaijan*
** Corresponding Author:email:depart15@iit.ab.az*
Note:This paper has been presented at the International Conference on Advanced Technology&Sciences (ICAT'14) held in Antalya (Turkey), August 12-15, 2014.

As a result of conducted researches, points reflecting the personal approach to recruitment of IT professionals emerged, which demonstrate themselves in different approaches to requirements set forth by the employer to the job applicant applying for the same position depending on the profile, activity direction, property type (government or non-government, joint etc.) of the organization.

This point emerges when a requirement indicated as obligatory by one employer for a specific position, can be evaluated as desired or even unimportant by another employer. Naturally, if a job applicant doesn't meet at least one indicator listed as obligatory for this specialty by the employer, his chances of getting accepted to the relevant position equals to zero.

Statistical results of the approach of 72 employers regarding meeting the indicators characterizing education and personal qualities for the specialty of programmer-engineer are presented in Table 1.

Accordingly, as semistructured, the problem of personnel selection is characterized by the following features:

- multifactorial and multicriteriality;
- criteria and indicators of qualitative and quantitative nature;
- the need to consider the experts views in the evaluation process;
- hierarchy rate criteria characterizing evaluated object, expressed in the fact that each top-level individual criterion is based on the aggregation of partial criteria;
- dependence on employer's requirements that define "portrait of the professional" to occupy particular position, at a real time.

Above listed characteristics of the issue defines the fuzziness of entry information, "loads" the issue to a fuzzy environment and this requires selection of an adequate formalism that considers the uncertainty of linguistic nature related to formalization of fuzziness of indicators and expert knowledge for modelling of the issue and evaluation of the alternatives. From this point of view, necessity for the use of fuzzy mathematical apparatus has emerged for solution of the recruitment issue.

Table 1. Result of employers requirements according to educational and personnel qualities criteria for proqrammer-engineer specialty

Character of employers' requirements / Indicators characterizing the employed person	obligatory (%)	desirable (%)	not required (%)
Education:			
Higher education diploma	68,11	25,02	6,87
Higher IT education diploma	30,58	51,43	17,99
Course certificates	5,64	31,97	62,39
Personal qualities			
Performance discipline	75,06	18,07	6,95
Initiative at work	23,63	55,52	20,85
Capability to pass on experience	13,9	56,91	29,19
Team work capability	34,67	29,19	36,14
Analytical thinking	17,99	50,04	31,97

3. Task Description

Current article proposes an approaching that enables to consider the individual requirements of the employers. Thus, we are proposing the approach that enables the selection of the best job applicant among all job applicants considering the individual requirement of the employer regarding meeting the general criteria indicators in order to be hired for specific IT specialties.

Thus, let's consider that $X = \{x_1, x_2, \ldots, x_n\} = \{x_i, i = \overline{1,n}\}$ – is a set of job applicants – alternatives the best of which must be selected; $K = \{K_1, K_2, \ldots, K_m\} = \{K_j, j = \overline{1,m}\}$ – is a set of criteria inherent to alternatives and the set is defined by knowledge, capability and personal qualities of job applicants. In this case, suitability of alternatives to criteria can be shown in two-dimensional matrix, whereas element of the matrix will be defined by membership functions reflecting the suitability level of x_i alternative to K_j criteria: $\varphi_{K_j}(x_i) : X \times K \to [0,1]$.

Here, $\varphi_{K_j}(x_i)$ – reflects the suitability level of xi alternative to *Kj* criteria. But these criteria are defined based on multiple indicators of different significance.

I.e. $K_j = \{k_{j1}, k_{j2}, \ldots, k_{jT}\} = \{k_{jt}, t = \overline{1,s}\}$.

Let's suppose,

1) $\{k_{jt}, t = \overline{1,s},\ j = \overline{1,m}\}$ membership function $\{\varphi_{k_{j1}}(x_i), \varphi_{k_{j2}}(x_i), \ldots, \varphi_{k_{js}}(x_i)\} = \{\varphi_{k_{jt}}(x_i), t = \overline{1,s},\ j = \overline{1,m}\}$ of $\{k_{jt}, t = \overline{1,s},\ j = \overline{1,m}\}$ alternatives to criteria indicators is known (supply base);

2) Evaluation of the decision making person (DMP) regarding obligation (O), desirability (D) and unimportance (U) of meeting $\{k_{jt}, t = \overline{1,s},\ j = \overline{1,m}\}$ criteria indicators for occupation of a specific position is known (requirement base).

Objective of the issue is to select the best alternative from the supply basis in accordance with demand basis for occupation of a specific vacancy or make a ranked list of alternatives from best to worst: $X : K^* \to X^*$. Hereby, X – is the set of primary alternatives, K^* – is the set of indicators marked with obligation (O), desirability (D) and unimportance (U), X^* – is the ranked list of selected alternatives in accordance with demand.

4. Issue solution

4.1. Modelling of the Demand Basis

Employer's criteria indicators $\{k_{jt}, t = \overline{1,T},\ j = \overline{1,m}\}$ for occupation of a specific vacancy are divided into three groups as obligatory (O), desirable (D) and unimportant (U) and form relevant sets: {O}, {D}, {U}.

Let's note that, $\{O\} \cap \{D\} \cap \{U\} = \varnothing$ and $\{O\} \cup \{D\} \cup \{U\} = \{k_{jt}, t = \overline{1,s},\ j = \overline{1,m}\}$, i.e. these sets do not have a common element, any $k_{jt} \in K_j \in K$ element can belong to only one of these sets. Following possible situations – scenarios can happen depending on distribution of $\{k_{jt}, t = \overline{1,s},\ j = \overline{1,m}\}$ criteria indicators among {O}, {D}, {U} sets.

Scenario 1. All indicators defining K_j criteria are obligatory: $k_{jt} \in \{O\}, t = \overline{1,s}$;

Scenario 2. A part of indicators defining K_j criteria are obligatory, another part is unimportant: $k_{jt} \in \{O\} \cup \{U\}, t = \overline{1,s}$;

Scenario 3. All indicators defining K_j criteria are desirable: $k_{jt} \in \{D\}, t = \overline{1,s}$;

Scenario 4. A part of indicators defining K_j criteria are desirable, another part is unimportant: $k_{jt} \in \{D\} \cup \{U\}, t = \overline{1,s}$;

Scenario 5. A part of indicators defining K_j criteria are obligatory, another part is desirable: $k_{jt} \in \{O\} \cup \{D\}, t = \overline{1,s}$;

Scenario 6. A part of indicators defining *Kj* criteria are obligatory, another part is desirable and a third part is unimportant: $k_{jt} \in \{O\} \cup \{D\} \cup \{U\}, t = \overline{1,s}$;

Scenario 7. All indicators defining *Kj* criteria are unimportant: $k_{jt} \in \{U\}, t = \overline{1,s}$.

(Let's note that, scenario 1 and 3 did not emerge during research and scenario 6 was the most common scenario).

4.2. Formation of the Supply Basis

Mathematical formalization of criteria must be carried out in order to find the membership function of

$\{k_{jt}, t=\overline{1,s},\ j=\overline{1,m}\}$

criteria indicators to alternatives.

K_1, K_2, K_3 are exact criteria and relevance of the job applicant to these criteria is determined in a formal order, based on the documentation submitted by the applicant.

An Indistinctness and quality characteristic, and support of expert knowledge during the definition of K_4, K_5, K_6 criteria, make it necessary to use fuzzy mathematical logic methods that enable to form the linguistic phrases of the natural language [18]. To that effect, it is necessary to develop mathematical formalization of criteria for realization of supply base, and the mechanism of turning the linguistic phrases regarding the level of satisfaction of criteria into a fuzzy value defined in the ─ [0,1] interval.

4.3. Mathematical formalization of Criteria

A criteria indicators scale is selected in order to determine the membership function – fuzzy value of the alternative criteria indicators, i.e. each criteria indicator is divided into rating levels in accordance with quality levels (excellent, good, acceptable, poor etc) of the relevant linguistic phrases of the natural language.

After performing of each criteria factor, appropriation of a fuzzy value from the fuzzy set to a linguistic rating level selected for it must be performed (Table 2).

Table 2. Mathematical formalization of "work experience in specialty"

Quality rating of "Work experience in specialty" indicator	Linguistic rating	fuzzy subset, set in [0, 1] interval
1) Has three or more years work experience in specialty	excellent	[0,98-1]
2) Has 1 to 3 years work experience in specialty	good	[0,8-0,97]
3) Has 6 months to 1 year work experience in specialty	acceptable	[0,5-0,79]
4) Has less than half a year work experience in specialty	poor	[0,1-0,49]

Final – collective fuzzy value determined by the experts based on individual fuzzy values can be defined in following ways:

1) by intersection of fuzzy sets;
2) by connection of fuzzy sets;
3) by making an agreed selection on fuzzy sets.

Based on the last approach, individual evaluation of the "superior" expert with special creativity is considered as the collective value. Such expert must choose such a membership value out of all individual membership values defined by experts as a collective membership value at each point of the possible alternatives space, that in general situation, it must differ from remote values in collective and hold a determined "middle" position.

Thus, a "supply basis" is formed by finding a

$$\{\varphi_{k_{j1}}(x_i), \varphi_{k_{j2}}(x_i), \dots, \varphi_{k_{jT}}(x_i)\} = \{\varphi_{k_{jt}}(x_i), t=\overline{1,s},\ j=\overline{1,m}\}$$

membership function based on how alternatives meet $\{k_{jt}, t=\overline{1,s},\ j=\overline{1,m}\}$ criteria indicators of alternatives.

5. Evaluation of alternatives

Evaluation of alternatives based on proposed indicators is carried out in three stages.

In the first stage, suitability of the job applicant to relevant requirements of the employer on indicators of K_1, K_2, K_3 criteria determined based on documents submitted by the job applicant. In the second stage, evaluation of alternative based on K_4, K_5, K_6 criteria is carried out. Definition of membership function of the alternative to these criteria is realized through a scenario relevant to evaluation of these criteria in the supply basis.

Claim 1. If a part of indicators defining $K_j = \{k_{jt}, t=\overline{1,s}\}$ (here $j=\overline{4,6}$) criteria (scenario 1, 2, 5, 6) is obligatory and the value of membership function of alternative to at least one of these indicators equals to 0, then the membership function of the alternative to the relevant criteria will also equal to 0.

Claim 2. $K_j = \{k_{jt}, t=\overline{1,s}\}$ (here $j=\overline{4,6}$) is only defined by desirable (or partly unimportant – scenario) indicators and the value of membership function of alternative to at least one of desirable indicators differs from 0, then the membership function of the alternative to the relevant criteria will also be different from 0.

Thus, membership function K_j, $j=\overline{4,6}$ of the alternative, depends on distribution of indicators characterizing it among {O}, {D}, {U} sets, scenarios.

Calculation of membership function of the alternative $K_j = \{k_{jt}, t=\overline{1,s}\}$ to the criteria, is based on membership function of the indicators characterizing the criteria and its "curve" i.e. their aggregation based on principal of their importance factor depicted in thus [19], following are proposed for calculation of membership function of the alternative to $K_j = \{k_{jt}, t=\overline{1,s}\}$ criteria:

1. Based on Scenario 1: Membership function of the alternative to criteria *Kj* is calculated using following equation.

$$\varphi_{K_j}(x_i) = \prod_{t=1}^{s} \left[\varphi_{k_{jt}}(x_i)\right]^{w_{jt}} \quad (1)$$

Here $\varphi_{k_{jt}}(x_i)$ – is the membership function of the job application to k_{jt} indicator, w_{jt} – is the importance factor of k_{jt} indicator. Let's note that,

$$\sum_{t=1}^{s} w_{jt} = 1, \quad t=\overline{1,s}$$

condition must be met for criteria indicators.

2. Based on Scenario 2: Suppose, g quantity of indicators defining *Kj* criteria have been evaluated as unimportant and naturally g<s. Then, the membership function formula of the alternative to *Kj* criteria (1) is defined based on s-g quantity of obligatory indicators.

3. Based on Scenario 3: Membership function of ith alternative to Kj criteria is calculated using

$$\varphi_{K_j}(x_i) = \sum_{t=1}^{s} w_{jt}\ \varphi_{k_{jt}}(x_i) \quad (2)$$

equation.

4. Based on Scenario 4, membership function of i^{th} alternative to *Kj* criteria is found only based on formula for indicators included in {D} set (2).

5. Based on Scenario 5, in order to find the , membership function of i^{th} alternative to *Kj* criteria, firstly the difference of membership function of its obligatory indicators from 0 is checked and if one of them equals to zero, then $\varphi_{K_j}(x_i)$ =0 is accepted, otherwise in accordance with formula (2), the value of membership function to *Kj* criteria is calculated. i.e.:

$$\varphi_{k_j}(x_i) = \begin{cases} 0, & if \quad \prod_{d=1}^{g} \varphi_{k_{js}}(x_i) = 0 \\ \sum_{t=1}^{s} w_{jt}\, \varphi_{k_{jt}}(x_i) & if \quad \prod_{d=1}^{g} \varphi_{k_j}(x_i) \neq 0. \end{cases} \quad (3)$$

Here, $k_{jd} \in \{M\}, d=\overline{1,g}$ ─ K_j is the obligatory indicators characterizing *Kj* criteria and naturally in this case g<s.

6. Based on Scenario 6, if S quantity of indicators of *Kj* is evaluated as unimportant, then it is possible to find the membership function of the alternative to this criterion by carrying out the

operational sequence relevant with formula (3) in accordance with *s-g* quantity of indicators.

7. Based on Scenario 7, during the definition of membership function of the alternative to *K*, (i.e. the value of the job applicant's chance to get the job), its membership function to *Kj* is not taken into consideration.

In the Third stage, the value of the job applicant's chance to get the job, i.e. $\varphi_K(x_i),\ i=\overline{1,n}$ must be defined. The value of membership function of alternative to K, is based on aggregation of its $\{\varphi_{K_j}(x_i),\ j=\overline{4,6}\}$ membership function to $K_j,\ j=\overline{4,6}$ criteria, i.e. the evaluation of the alternative's chance to get the job is defined based on $\varphi_K(x_i)=\sum_{j=4}^{6} w_j\ \varphi_{K_j}(x_i)$ formula [19], [20].

6. Use of information about importance of the criteria

This point is one of the problems emerging in the solution of personnel management problems and obtaining of such information gives opportunity to eliminate multicriterionness and to bring this problem to one-criterion problem. In this case global criterion is defined as

$$K_Q=\sum_{j=1}^{m} w_j K_j$$

And here K_j ─ is criterion characterizing estimated object (j=1, 2,…, m), w_j ─ is called weight of criterion K_j or importance factor [21]. For importance factor of the criterion the following condition is foreseen:

$$0\le w_j\le 1;\quad \sum_{j=1}^{m} w_j=1 \qquad (4)$$

The idea of unification is based on the expressions of the person who expresses the opinion about importance of criteria (expert, person who makes a decision) or on determination of appropriate evaluation grade determined to reflect value of considered criterion(in other case refer to 1-100 point scale) and further normalization within condition (1) of this value. On the basis of the obtained information for today preparation of methods for determining of criteria importance factors is one of the points the attention is attracted to in the sphere of multicriterion problems solution [21], [22].

Information about mutual importance, significance of the criteria can be referred by the experts can be:

─ expressed by the linguistic expressions representing mutual relative advantage (or weak points) and their pair comparison;

─ referred to the establishing of appropriate grade to reflect assessment value of the considered criterion against the background of criteria defining any global factor.

In first case to display mutual relative advantage of the criteria the linguistic expressions of the type given below are used:

- criterion K_1 has a weak advantage over criterion K_2
- criterion K_2 has rather more advantage over criterion K_1 and etc.

Such linguistic expressions for degree of mutual relative advantage of compared criteria are estimated by 9-point Saati's table (Table 3) [23].

If number of criteria equals to *n* then by defining of n-1 ratio of pair criteria comparison it is possible to make a matrix of mutual relative relations [23], [24].

Table 3. Defining of relative importance factors of pair comparison on the basis of quality estimations

Importance intensity	Qualitative (linguistic) estimation
1	Criterion K_1 has no advantage over K_2
3	Criterion K_1 has weak advantage over K_2
5	Criterion K_1 has essential advantage over K_2
7	Criterion K_1 has evident advantage over K_2
9	Criterion K_1 has absolute advantage over K_2
2,4,6,8	Intermediate estimations between neighboring estimation

After all matrix elements are defined private vector (w_i^*) is to be found. For this purpose radical of *n*-power of matrix line edge (n is measure of comparison matrix) should be defined and after they are normalized importance factor w_i of appropriate elements is calculated.

$$w_i^*=\sqrt[m]{K_{i1}\times K_{i2}\times\ldots\times K_{im}}$$

$$w_i=\frac{w_i^*}{\sum_{i=1}^{m} w_i^*} \qquad (5)$$

It must be noted that importance factors identified by means of formula (5) condition (4) is being checked up.

In the second case information about the importance, significance against the background of common criteria reflects value of any criterion.

In such case it is more advantageous to use method of importance factor on the basis of 10-point system of expert estimation of the criteria [24].

7. Detection of contradictions in the expressions of pair comparison about criteria importance

It must be noted that usually in multicriteria tasks multiple number of criteria and criteria indicted lead to the contradictions of expert expressions reflecting their pair comparison made by expert group members.

Thus before the application of criteria importance factor found by formula (5) in appropriate way one of the primary task is to identify if contradictory information (expert knowledge) used for their pair comparison is available. For this purpose maximal private value $\lambda_{\max}$, consent index and consent relation must be calculated.

Calculation of maximal private value $\lambda_{\max}$ is implemented by the pair comparison matrix as follows: each column of expressions is summarized, then sum of the first one is multiplied to the quantity of the first component of normalized priority vector, and sum of the second column is multiplied with second one and etc., then all obtained numbers are added. I.e.,

$$\lambda_{\max}=\sum_{i=1}^{n}\left(\sum_{j=1}^{n} k_{ij}\times w_i\right)$$

The closer $\lambda_{\max}$ is to *n* (*n* – is a number of compared matrix elements), the more consent the result is.

Decline from consent may be expressed by the value $(\lambda_{\max}-n)/(n-1)$, that will be called consent index (consent index – *CI*).

CI is calculated by the following formula:

$$CI=(\lambda_{\max}-n)/(n-1).$$

If *CI* is divided into the number appropriate to the chance consent – *CC*, we obtain consent relation – *CR*.

According to [23] for matrix of the *n=3* size chance consent *CC=0,58;* for matrix of the *n=4* size *CC=0,90*; for *n=5* size *CC=1,12;* for *n=6 size CC=1,24* and etc.

Consent relation if identified by the following formula:

$$CR=CI/CR.$$

Consent rate is considered acceptable at $CR\le 0{,}1$. If consent rate if higher than 0,1, then expressions should be re-considered.

8. Defining of importance factor by pair comparison

In the Institute of Information Technologies of ANAS general criteria system of nominees employment problem on the IT-specialties has been created in the framework of building of personnel management intelligent system [3].

The list of criteria for recruitment as an IT professional, set forth by the employers for those wishing to be employed have been determined. Criteria are presented in 6 groups: criteria are presented as following, K_1 – age, K_2 – gender, K_3 – education, K_4 – personal qualities, K_5 – professional requirements in IT specialization, K_6 – additional capabilities. Each of these criteria is defined by multiple indicators that characterize them for instance K_4 – personal quality criteria is determined based on below indicators: k_{41} – performance discipline; k_{42} – initiative at work; k_{43} – capability to pass on experience; k_{44} – team work (communication) capability; k_{45}– analytical thinking. On the basis of the expressions said by the expert about theoretical importance of these shown criteria indices the given below (Table 4) relation matrix is created by using relational importance scale displayed in Table 3. While matrix is being compiled it is referred to the its diagonal, symmetric and transitive features. For instance because of evident superiority of criteria index k_{44} over criterion index k_{42} 5 is written in appropriate cell of the matrix, while in diagonally symmetric place cell 1/5 is noted.

Table 4. Comparison matrix personal quality criteria indicators

	k_{41}	k_{42}	k_{43}	k_{44}	k_{45}	Private vector (w_i^*)	Importance factor (w_i)
k_{41}	1	4	4	0,33	1	1,39	0,22
k_{42}	0,25	1	1	0,2	2	0,63	0,1
k_{43}	0,25	1	1	0,25	0,5	0,57	0,09
k_{44}	3	5	4	1	4	2,99	0,47
k_{45}	1	0,5	2	0,25	1	0,76	0,12
$\sum_{J=1}^{5} k_{4j}$	5,5	11,5	12	2,03	8,5		

After matrix has been compiled importance factors of the criteria are found by means of formula (5). In next step the availability of contracting features of used expert expressions is checked. For this purpose first of all $\lambda_{\max}$ is found.

$$\lambda_{\max} = \sum_{i=1}^{5}(\sum_{j=1}^{5} k_{ij} \times w_i) = 5{,}41.$$

consent index (CI) of the used expert expressions is defined.

$$CI = (\lambda_{\max} - n)/(n-1) = 0{,}102 .$$

If we consider chance consent to be CC= 1,12 for the 5-sized matrix then we can calculate consent relation – CR.

$$CR = CI / CR = 0{,}09 .$$

Consent relation was defined to be lower than 0,1 and it means there is no contradiction in the expressions used by the experts about criteria pair comparison and determined importance factor can be used in the realization of the task.

9. Conclusion

A system supporting the decision making system is used in ANAS Institute of Information Technology during the recruitment of IT specialized staff. The results of the survey with 101 specialist-experts specialized in IT field are used for the formation of the information base of the system. It is considered to use the results obtained from the realization of the system for decision making during regulation and administration of supply and demand in IT labour market.

Proposed solution method of the issue of recruitment of IT specialized staff is realized in Delphi 2010 programming system. The proposed methodology and decision support system is successfully used in various companies to support management decision making for the recruitment of IT professionals. The application of the system required to improve methodology towards the preferences and interests of IT professionals. Currently, the work is underway to develop a method for making trade-off decisions to deal with the preferences of employers, as well as IT professionals.

References

[1] M. H. Mammadova, Z. G. Jabrayılova, M. I. Manaflı, Monitoring of demands for Information Technology specialists, Baku, 2009. p. 199.

[2] M. Mamedova, F. Mamedzade, "Fuzzy multicriterion estimation of demand for IT-specialities", in Proc. International Conference "Problems of Cybernetics and Informatics (PCI─2010), Baku, Azerbajan 6─9 September ,2010, vol. 1, pp. 94–97.

[3] W. C. Borman, M. A. Hanson, J. W. Hedge, "Personnel selection", Annual review of psychology, 1997, pp. 299–337.

[4] T. Robertson, M. Smith, "Personnel selection", Journal of Occupational and Organizational Psychology, 2001, vol.74, Issue 4, pp.441–472.

[5] M. Dursun, E. Karsak, "A fuzzy MCDM approach for personnel selection", Expert Systems with Applications, 2010, v.37, pp.4324–4330.

[6] L. S. Chen, C. H. Cheng, "Selecting IS personnel use fuzzy GDSS based on metric distance method", European journal of operational research, 2005, vol.160, no 3, pp. 803–820.

[7] Z. Gungor, G. Serhadlıoglu, S. E. Kesen, "A fuzzy AHP approach to personnel selection problem". Applied Soft Computing, 9 (2009), pp. 641–649.

[8] Y. J. Wang, H. S. Lee, 'Generalizing TOPSIS for fuzzy multiple-criteria group decision-making", Computers and Mathematics with Applications, 2007, v.53, no.11, pp.1762–1772.

[9] A. Kelemenis, D. Askounis, "A new TOPSIS-based multi-criteria approach to personnel selections", Expert systems with applications, vol 37, pp. 4999–5008, 2010.

[10] M. H. Mammadova, Z. G. Jabrailova, S. M. Nobari, "Application of TOPSIS method in support of decisions made in staff management issues", in Proc. IV International Conference "Problems of Cybernetics and Informatics" (PCI-2012), Sept.12-14, 2012, Baku, vol. IV, pp.195-198.

[11] C. F. Chien, L. F. Chen, "Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry", Expert system with applications, 2008, vol 34(2), pp 280–290.

[12] P. C. Chen, "A Fuzzy Multiple criteria decision making model in employee recruitment", IJCSNS International Journal of Computer Science and Network Security, 2009, vol. 9, no.7, 113-117.

[13] M. S. Mehrabad, M. F. Brojeny, "The development of on expert system for effective selection and appointment of the jobs applicants in human resource management", Computer and Industrial engineering, 2007, vol 53, pp.306-312.

[14] M. Nussbaum, M. Singer, R. Rosas, M. Castillo, E. Flies, R. Lara and R. Sommers, "Decision support system for conflict

diagnosis in personnel selection," Information & Management, Vol. 36, Issue 1, pp. 55-62, 1999.
[15] O. I. Larichev, M. Sternin, "Decision support system of multi-objective problem of assignment", Information systems and processes, 1998, Vol. 3, pp.10-16.
[16] R. S. Hooper, T. P. Galvin, R. A. Kolmer, J. Liebowitz, "Use of an expert system in a personnel selection process", Expert systems with Applications, 1998, vol 14(4), pp.425-437.
[17] M. H. Mamedova, Z. G. Jabrayilova, S. M. Nobari, "Methodology for solution of personnel employment task for multi-scenario analyze", Problems of information technology, Baku, no. 2, pp. 24–35, 2010.
[18] L.A.Zadeh, "The Concept of a Linguistic variable and its Application to Approximate Reasoning-I", Information Sciences, no. 8, pp. 199–249, 1975.
[19] J. V. Neumann, O. Morgenstern, Theory of games and economic behavior, One of Princeton University press's Notable Centenary Titles, 2007, 776 pp.
[20] P. V. Sevestyanov, L. G. Dimova, M. Kaptur, A. V. Zenkova, "Methodology of multi-criterion hierarchic estimation of quality in uncertainty conditions", Information technologies, Moskau, no.9, pp. 10–13, 2001.
[21] O. I. Larichev, Theory and Methods of Decision Making, Logos, Moscow, 2000, p. 296 (in Russian).
[22] V. D. Nogin, Decision making at muptiple criteria, St. Petersburg, p.103, 2007. (in Russian).
[23] T. L. Saaty, The Analytic Hierarchy Process, Mc. Graw-Hill, 1980, 267p.
[24] Z. G. Jabrayilova, S. N. Nobari, "Processing methods of information about the importance of the criteria in the solution of personnel management problems and contradiction detection", Problems of information technology, Baku, 2011, no 1, pp. 57–66.

International Journal of
Intelligent Systems and
Applications in Engineering
ISSN:2147-6799 www.atscience.org/IJISAE
Advanced Technology and Science
Original Research Paper

Particle Swarm Optimization Based Approach for Location Area Planning in Cellular Networks

Mays S. Algebary*[1], Issmail M. Ellabib[1], Ahmed B. Elwhishi[2]

Received 27th September 2014, Accepted 22th December 2014

Abstract: Location area planning problem plays an important role in cellular networks because of the trade-off caused by paging and registration signalling (i.e., location update). Compromising between the location update and the paging costs is essential in order to improve the performance of the network. The trade-off between these two factors can be optimized in such a way that the total cost of paging and location update can be minimized along with the link cost. Due to the complexity of this problem, meta-heuristic techniques are often used for analysing and solving practical sized instances. In this paper, we propose an approach to solve the LA planning problem based on the Particle Swarm Optimization (PSO) algorithm. The performance of the approach is investigated and evaluated with respect to the solution quality on a range of problem instances. Moreover, experimental work demonstrated the performance comparison in terms of different degree of mobility, paging load, call traffic load, and TRX load. The performance of the proposed approach outperform other existing meta-heuristic based approaches for the most problem instances.

Keywords: *Particle Swarm Optimization, Simulated Annealing Optimization, Ant Colony Optimization, Location Management in Cellular Networks, Swarm Intelligence.*

1. Introduction

The main purpose of the network planning is to manage the resources of the mobile cellular networks in order to reduce the cost of operation and meet the quality of service. [3]. There are two basic operations involved with location management; location update and paging. The goal is to partition the network into a given number of location areas such that the total cost of paging and location update is maintained at its minimum. In particular, upon the arrival of a mobile-terminated call, the system tries to find the mobile terminal by searching for it among a set of Base Transceiver Stations (BTSs) over the current region of the mobile. This search is called paging, and the set of Cells of BTSs in which a mobile is paged is called the Location Area (LA). At each LA boundary crossing, mobile terminals register (i.e., update) their new location through signalling in order to update the location management databases. Finding appropriate size or the number of LAs is essential for reducing the costs of paging and location updates signalling [3].

Finding the optimal number of location areas and the corresponding configuration of the partitioned network has motivated many researchers as it is a difficult combinatorial optimization problem which is classified as an NP- hard problem [3], [6]. The previously reported studies divided the whole problem into two sub-problems, the cell-to-switch assignment problem, and the LA planning problem [2]. Many approaches have been introduced to solve the cell-to-switch assignment problem in the literature, such as Genetic Algorithm [5], Evolutionary Algorithm [4], Ant Colony System [8], and Particle Swarm Optimization [11]. However, these approaches are based on the static location management scheme, where the service area is divided into fixed LAs sizes, and the users in a given region are assigned to the same LA regardless of their characteristics. For the economic feasibility of any communication system, the good design method should optimize the network cost, while considering some factors such as traffic, bandwidth, and capacity. The weakness of the above approaches is that the paging cost is not considered. Moreover, it is assumed that each switch manages only one LA that is equal the size of the cells belonging to that MSC, which further degrades the quality of the solution. On the other hand, most of recent personal mobile network systems, including the GSM system, are employing the zone based scheme due to high mobility and increase of subscribers [6]. In this scheme, the service area is divided into groups of cells forming LAs. The mobile terminals update their locations only when they leave their current LAs and enter new LAs. The optimization of this scheme has not been widely studied, unlike that of static based scheme. Only few studies have addressed the LA planning problem for the zone based scheme such as Simulated Annealing Algorithm [6] and Ant Colony System [1]. In this paper the Particle Swarm Optimization (PSO) algorithm is adapted to provide a solution to the LA planning problem for the zone based scheme, since most of the existing personal mobile networks use the zone based scheme in practice and include all its realistic objectives and constraints.

The rest of the paper is organized as follows: Section.2 presents the proposed Particle Swarm Optimization (PSO) approach for the LA planning problem. Section.3 provides a set of computational experiments for the analysis and performance comparison. Finally, some concluding remarks are provided in Section.4.

2. The Proposed Particle Swarm Optimization (PSO) Approach

LA planning problem is defined as follows: the service area is divided into a number of Location Areas (LAs) and each LA

[1] *Tripoli University, Computer Engineering Department, Tripoli, Libya*
[2] *Faculty of Electronics Technology, Tripoli, Libya*
* *Mays S. Algebary: Email: mais.muhsen@yahoo.com*
Note: This paper has been presented at the International Conference on Advanced Technology&Sciences (ICAT'14) held in Antalya (Turkey), August 12-15, 2014.

consists of a number of Base Transceiver Station (BTSs). The problem arises from the trade-off between the location update and the paging costs, along with the consideration of the link cost which represents the distance between BTSs, Base Station Controllers (BSCs), and Mobile Service Switching Centres (MSCs). The challenge is to assign BTSs to BSCs, BSCs to MSCs, and BTSs to LAs considering the assignment costs such that the total cost is minimized and a set of realistic constraints must be satisfied. For more details on the mathematical problem formulation, the reader may refer to [1] and [6].

2.1. PSO algorithm

Particle Swarm Optimization (PSO) is a population based search algorithm inspired by bird flocking and fish schooling originally designed and introduced by (Kennedy et al; 1997) [9]. In contrast to evolutionary computation paradigms such as Genetic Algorithm, a swarm is similar to a population, while a particle is similar to an individual. Typically, the particles fly through a multidimensional search space in which the position of each particle is adjusted according to its own experience and the experience of its neighbours. In binary (discrete) version, each particle is composed of D elements which indicate a potential solution [9]. The appropriateness of the solution is evaluated by a fitness function. Each particle is considered as a position in a D-dimensional space and each element of a particle position can take the binary value of 0 or 1 in which 1 means "included" and 0 means "not included". Each element can change from 0 to 1 and vice versa. Also, each particle has a D-dimensional velocity vector, the elements of which are in range [Vmin, Vmax]. Velocities are defined in terms of probabilities that a bit will be in one state or the other. At the beginning of the algorithm, a number of particles and their velocity vectors are generated randomly. Then, the velocity (V) and position vector (X) are updated iteratively using (Equation.1) and (Equation.2), respectively,

$$V_i^{t+1}(j) = W\,V_i^t(j) + C_1 r_1\left(pbest_i^t(j) - X_i^t(j)\right) + C_2 r_2\left(nbest_i^t(j) - X_i^t(j)\right) \quad (1)$$

$$X_i^{t+1}(j) = \begin{cases} 1 & if\ \frac{1}{1+exp(-V_i^{t+1}(j))} > r_{ij} \\ 0 & otherwise \end{cases} \quad (2)$$

$X_i^t(j)$ is the position of j^{th} element of i^{th} particle in t^{th} iteration. $V_i^t(j)$ is the velocity of j^{th} element of the i^{th} particle in t^{th} iteration. The *pbest* is the current best position of the particle, and *nbest* is the best position found so far. C_1 and C_2 are positive acceleration constants which control the influence of *pbest* and *nbest* on the search process. r_1 and r_2 are random values in the range [0, 1] that are sampled from a uniform distribution. W is called inertia weight, which is introduced to control the exploration and exploitation abilities of the swarm. r_{ij} of (Equation.2) is random number on the range [0, 1]. The advantage of the PSO algorithm is the combination of local search methods (through self-experience) with global search methods (through neighbouring experience). However, the set of parameters in (Equation.1) need to be tuned sensibly in order to balance exploration and exploitation for the search of high quality solution. For more details on the PSO algorithm, the reader may refer to [7] and [9].

2.2. Adaption of PSO algorithm

In order to adapt PSO algorithm for solving the LA planning problem, we represent the problem in the form of three following connections: the BTS-BSC connection, BSC-MSC connection, and BTS-LA connection such that each particle can be encoded as three $m \times n$ matrices, which called the position matrices of the particle. m represents the number of BTSs in the matrix of BTS-BSC connection and BTS-LA connection, and the number of BSCs in the matrix of BSC-MSC connection. n represents the number of BSCs in the matrix of BTS-BSC connection, the number of MSCs in the matrix of BSC-MSC connection, and the number of LAs in BTS-LA connection. Each position matrix in each particle has the following properties:

- All the elements of the position matrices have either the value of 0 or 1.
- In each row of these matrices only one element is 1 and others are 0.

In position matrix 1, each column represents a BSC and each row represents allocated BTS in a particular BSC. In position matrix 2 each column represents MSC and each row represents BSC allocated in a particular MSC. In position matrix 3 each column represents LA and each row represents BTS allocated in a particular LA. For instance, (Table.1) illustrates the position matrix 1 (BTS-BSC connection) for one particle; the position matrix 1 illustrates that BTS1 is assigned to BSC1, BTS2 is assigned to BSC2, BTS3 is assigned to BSC2, etc. The same interpretation is applied for the other matrices

Table 1. Position matrix 1 for a single particle

	BSC1	BSC2	BSC3	BSC4
BTS1	1	0	0	0
BTS2	0	1	0	0
BTS3	0	1	0	0
BTS4	0	0	1	0
BTS5	0	0	0	1

Velocity of each particle is considered as a m×n matrix whose elements are in range [V*min*,V*max*]. In other words, if $V_k$is the velocity matrix of k^{th}particle, then:

$$V_k^t(i,j) \in [V_{min}, V_{max}] \quad (\forall i,j), i \in \{1..m\}, j \in \{1..n\} \quad (3)$$

*pbes*t and *nbest* are $m \times n$ position matrices and their elements are 0 or 1. $pbest_k$ represents the best position that k^{th} particle has visited since the first iteration and $nbest_k$ represents the best position that k^{th} particle and its neighbors have visited from the beginning of the algorithm. For updating *nbest* in each iteration, *pbests* are used so that if the fitness value of *pbest* is greater than *nbest*, then *nbest* is replaced with *pbest*. (Equation.4) is applied for updating the velocity matrix and (Equation.5) is applied to update the position matrix of each particle.

$$V_k^{(t+1)}(i,j) = W.V_k^t(i,j) + c_1 r_1\left(pbest_k^t(i,j) - X_k^t(i,j)\right) + c_2 r_2\left(nbest_k^t(i,j) - X_k^t(i,j)\right) \quad (4)$$

$$X_k^{(t+1)}(i,j) = \begin{cases} 1 & if\ \left(V_k^{(t+1)}(i,j) = \max\left\{V_k^{(t+1)}(i,j)\right\}\right) \\ 0 & otherwise \end{cases} \quad (5)$$

Where, $V_k^t(i,j)$ denotes the element in i^{th} row and j^{th} column of the k^{th} velocity matrix in t^{th} iteration of the algorithm, and $X_k^t(i,j)$ denotes the element in i^{th} row and j^{th} column of the

k^{th} position matrix in t^{th} iteration. (Equation.5) illustrates that in each column of position matrix a value 1 is assigned to the element whose corresponding element in velocity matrix has the max value in its corresponding column. If there is more than one element in the column of velocity matrix with max value, then one of these elements is selected randomly and 1 assigned to its corresponding element in the position matrix. (Figure.1) illustrates the basic steps of the adapted structure of the PSO algorithm for solving the LA planning problem.

```
- Initialize particles with random positions.
 - Initialize each particle with random velocity.
 - Initialize pbest and nbest.
While iterations<= Max_of _iterations
For each particle
    Calculate the fitness value from the obj. function.
    If current fitness value >fitness value of the pbest
                Then update pbest
End for
-Update nbest from pbest.
-Apply assignment criteria (determine the assignment type).
- For each particle
        - Update the particle velocity using Equation.4
        - Update the particle position using Equation.5
- End For
End while
Return the solution from the nbest.
```

Figure 1. The adapted structure of PSO algorithm.

2.3. Neighbourhood Structure (Assignment Types)

In this paper, three types of assignments are introduced for the connection of BTSs, BSCs, and MSCs. The feasible solution can be generated by any of the three types of assignment and each assignment affects the network structure in different ways.

- BTS to BSC Assignment: In this type, the BTS-BSC connection or position matrix of the particle is considered. (Equation.4) will be used to update the velocity matrix related to BTS-BSC connection, and the BTS-BSC position matrix is consequently updated by (Equation.5). Thereafter the new load on each BSC is updated and all the constraints affected by this BTS-BSC new connection are checked. If these constraints are violated then this particle is marked at the end of the iteration as invalid (infeasible solution) otherwise it is marked as valid. Finally, for each MSC starting with one LA, BTSs are assigned to an LA. If LA capacity (i.e., paging capacity of BTSs) reaches to its limit, then a new LA is created and remaining BTSs started to be assigned to that LA. Here, the aim is to create the minimum number of LAs for each MSC. The reason of marking the particles at the end of each iteration as valid or invalid is that; the infeasible solutions that is found with invalid particles which means they did violate the constraints, may eventually lead to better positions/solutions, or may not. Thus, it is might be better to use some of these invalid particle in each iteration. In order to control the number of invalid particles in each iteration so they will not be the majority and eventually affect the quality of the solution with iterations, a mechanism of accepting a predefined ratio of invalid particle to the valid once in each iteration has been used. Thus, number of invalid particles can be controlled in each iteration.
- BSC to MSC Assignment: In this assignment type, velocity and position matrices are updated on the BSC-MSC connection without affecting the BTS-BSC connection. First, the BSC-MSC velocity matrix is updated using (Equation.4) and then the BSC-MSC position matrix is updated using (Equation.5). Due to the new connections caused by (Equation.5), the feasibility of the capacity and proximity constraints of that new BSC-MSC connection must be checked. Particles that violate the constraints will be marked as invalid at the end of the iteration for the same purpose illustrated in BTS-BSC assignment type. Finally, for each MSC starting with one LA, BTSs are assigned to an LA. If LA capacity (paging capacity of BTSs) reaches to its limit, then a new LA is created and remaining BTSs started to be assigned to that LA.
- BTS to LA Assignment: In this assignment type, the particle changes the BTS-LA assignment without affecting the BTS-BSC connection i.e., the particles search for all the LAs residing within the same BSC. The assignment is done by updating the velocity matrix of BTS-LA connection using (Equatio.4), then the BTS-LA position matrix is updated using (Equation.5). Subsequently, the capacity constraint of BTS-LA connection is checked, in case this constraint is violated then this particle will be marked as invalid, the position/solution will not be accepted, and the particle will continue with its old connection/position (solution).

3. Computational Experiments and Analysis

Several experiments are conducted for parameter setting and performance evaluation of the PSO based approach. An extensive performance study is carried out to evaluate the effectiveness of the proposed approach for the LA planning problem. The proposed PSO is compared to the Simulated Annealing (SA) and the Ant Colony System (ACS) based approaches. For more details on SA and ACS based approaches for solving LA problem, the reader may refer to [6] and [1], respectively. The proposed approach and the other two meta-heuristic based approaches are implemented using visual C++ and executed under Windows8 operating system, with Intel® core ™ i5-3210M 2.50 GHz CPU and 12 GB RAM. The computation time of the approaches varies with the problem size. The data sets vary according to different patterns (high, moderated, and low) of mobility, paging rate, call traffic load, and TRX load.

3.1. Parameter Setting

The velocity update equation (Equation.4) depends on a number of parameters that need to be determined in order to provide high quality solution. In this paper, the PSO approach is applied to a typical moderately size sample network that consists of 398 BTSs, 2 BSCs, and 1 MSC for the parameter setting experiments. The preliminary experiments are performed starting with an initial setting of the parameters based on values previously reported in the literature.

The PSO tends to have more global search ability at the beginning of the run while having more local search ability near the end of the run in order to refine a candidate solution.

The inertia weight (W) has some valuable information about previously explored directions. It governs how much of the previous velocity should be retained from the previous time step [10]. The best found starting value of W is 1.2 and gradually declined towards 0 during the search.

The acceleration parameters C_1 and C_2 denote the direction of the

particle towards optimal positions. They represent "cognitive" and "social" component, respectively. They affect how much the particle's local best and the global best influence its movement. From the experimental results it found that giving equal chances to exploration and exploitation didn't improve the cost as good as when giving different values for C_1 and C_2. It has found that choosing a larger cognitive parameter,C_1 than a social parameter, C_2, ($i.e.$, C_1=1.5 , C_2=0.5) gives the best result.

To ensure convergence of the heuristic, every velocity vector is bounded component-wise by minimum and maximum values [Vmin, Vmax]. These parameters are proved to be crucial, because the maximum velocity Vmax serves as a constraint to control the global exploration and exploitation ability of the particle swarm. The best result is achieved when Vmax and Vmin are equal to 50 and 0.05, respectively.

The type of assignment is chosen according to some probabilities assigned for each type of assignment. The best quality solutions were obtained, when the probability assignment of the BTS-BSC is 0.2, BSC-MSC assignment is 0.5, and BTS-LA assignment is 0.3.

The quality of solution is affected by the number of particles participating in the search process. More particles indicate a more cooperative interaction. To determine the appropriate number of particles, the results obtained for different numbers of particles, and it has found that the best number of particle is 30.

The parameters values have been tested and verified with different problem instances in terms of network cost. (Figure.2) depicts the cost of solution found at different stages of a typical run of PSO based approach for a sample network.

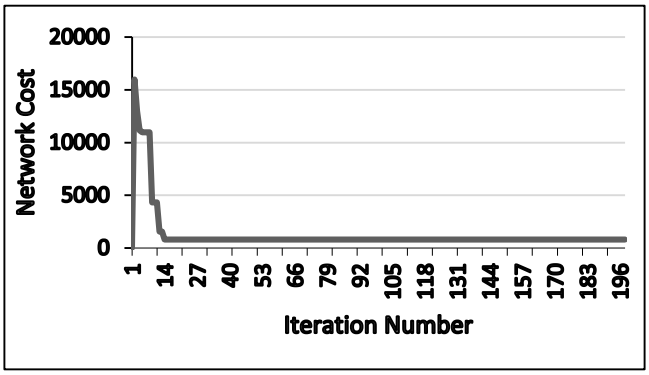

Figure 2. A typical run of the PSO based approach on a sample network.

3.2. Performance Evaluation and Comparison

To demonstrate the potential of applying the PSO algorithm for solving the LA planning, the PSO approach is compared with other existing meta-heuristic approaches. The PSO is compared with the SA and the ACS approaches on different network data sets. The network consists of 203 BTSs, 6 BSCs, and 3 MSCs. The parameters values of the SA and the ACS are set as reported in [6] and [1], respectively. To achieve the feasibility of the data generated for the network, the generation of the network loads is done according to the average values measured on the GSM network [6]. The three approaches are evaluated on the basis of the average of 30 independent runs for each instance.

Results denoted with (*) represent the results of SA and ACS as reported in [6] and [1], respectively. As shown in (Table.2), the PSO obtained the best solution for all datasets.

In what follow, the performance of the three approaches is investigated and evaluated on the four data sets with low and high values of the network factors such as mobility, paging rate, call traffic load, and TRX load. Each network factor value is scaled up (denoted by H*i*) and scaled down (denoted by L*i*) for data set *i*.

Table 2. Performance comparison between different approaches

Data Set	SA	ACS	PSO
1*	9880	9685	7687
2	7453	6866	6256
3	8951	8148	6058
4	6529	5997	5332

3.3. Effect of Mobility

To achieve a network with high mobility, the value of the handover rate of the network is scaled up by 70%. For low mobility, the same values of the handover rate is scaled down (divided by two). The other factor values remain unchanged. (Figure.3) demonstrates the performance for the three approaches (SA, ACS, and PSO) on four data sets for high mobility (H1, H2, H3, and H4) and for low mobility (L1, L2, L3, and L4).

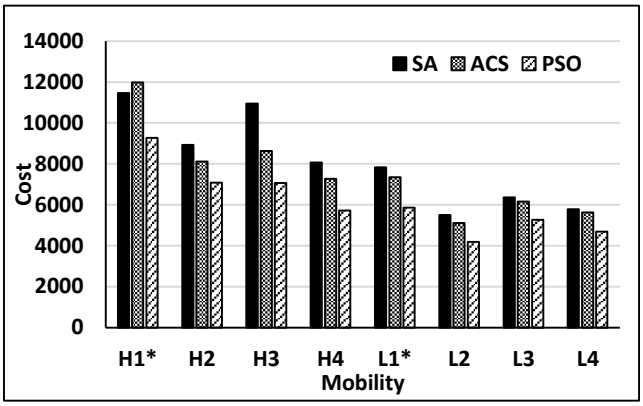

Figure 3. Performance comparison between different approaches with different mobility.

In (Figure.3), the obtained costs have increased as a result of the high crossing rate between the neighboring cells. When the handover rate increases, the position update equation (Equation.5) is affected by the handover constraints and the obtained costs are slightly higher, but still the PSO has the best solution compared to SA, and ACS. For network with a low mobility, the obtained cost is low compared to that of the high mobility case.

3.4. Effect of Paging Load

The network with high and low paging load are generated to examine the performance of the PSO algorithm. The paging load results from the number of mobile terminated calls, generated in a unit time. A high paging load means the number of mobile terminated calls is high, and a low paging load means the number of mobile terminated calls is low. A high paging load is obtained by increasing the values of the paging load for each BTS by 70%. The low paging load is obtained by scaling down the values of the paging load by 50%. The other values remain unchanged. (Figure.4) demonstrates the performance for the three approaches (SA, ACS, and PSO) on the four data sets for high paging rate (H1, H2, H3, and H4) and for low paging rate (L1, L2, L3, and L4).For the high paging load and low paging load, (Figure.4) shows the PSO presents better results compared to SA and ACS.

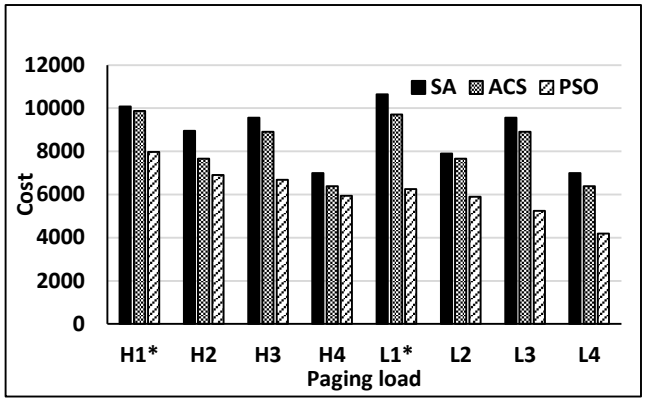

Figure 4. Performance comparison between different approaches with different paging load.

3.5. Effect of Call Traffic Load

A high traffic load is generated by scaling up the values of the call traffic load by 70%, and the low call traffic load is generated by dividing the call traffic load by two. The other factor values remain unchanged. (Figure.5) demonstrates the performance for the three approaches (SA, ACS, and PSO) on the four data sets for high call traffic load (H1, H2, H3, and H4) and for low call traffic load (L1, L2, L3, and L4).

(Figure.5) illustrates the performance of the PSO over the ACS and the SA. If we compare the results in (Figure.5) with the results in (Table.2), it noticed that these results are almost the same for high and low call traffic load for all data sets. This can be interpreted as the call traffic is a constraint which can be handled as long as the capacity of BSC and MSC is enough to handle those loads, it doesn't affect the total cost in direct way. In the meanwhile it would make the search space more tightened for the particles, which would eventually affect the quality of the solution and the constraint might be violated due to the restricted search space.

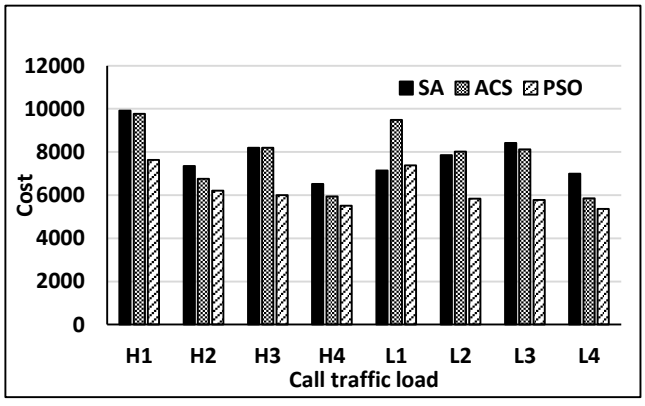

Figure 5. Performance comparison between different approaches with different call traffic load.

3.6. Effect of TRX

High and low TRX are generated by scaling up the moderate values by 70%, and down by 50%, respectively. The other factor values remain unchanged.

(Figure.6) demonstrates the performance for the three approaches (SA, ACS, and PSO) on the four data sets for high TRX (H1, H2, H3, and H4) and for low TRX (L1, L2, L3, and L4). (Figure.6) illustrates the performance of the PSO over the ACS and the SA. It noticed that solutions of high TRX is slightly better than low TRX. The reason is that the capacity constraints (in the case of the low TRX) are relaxed. As a result, the search space becomes larger and finding better solutions becomes harder.

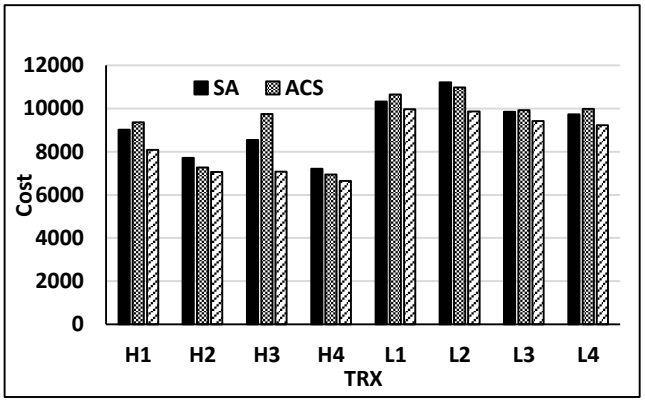

Figure 6. Performance comparison between different approaches with different TRX.

4. Conclusions

The most important benefit of optimized LA planning is preventing unnecessary radio resource usage that can instead be allocated for the communication of the subscribers. The network resources can be utilized more efficiently, and the network construction costs can be reduced by applying meta-heuristic to cellular network planning problems which leads to the proper use of limited radio resources. In this paper, the particle swarm optimization algorithm is adapted to solve the LA planning problem. The main goal is to develop an approach that can solve the LA planning problem more efficiently with particular emphasis on the tuning of the PSO parameters and the resource assignment criteria.

The potential improvement has been accomplished through the design and the analysis of PSO approach. The experimental results have shown that the PSO approach outperforms other meta-heuristic approaches on different data sets with different network loads of practical cellular network. However, further improvement to enhance the proposed approach is still needed. Of particular importance is investigation on the parameter sensitivity.

Acknowledgment

This research is partially supported by the Libyan National Agency for Scientific Research (NASR). This support is gratefully acknowledged.

References

[1] A. Elwhishi, I. Ellabib, and I. El-Feghi,"Ant Colony Optimization for Location Area Planning Problem in Cellular Networks," In the 9th International Arab Conference on Information Technology, Hammamet-Tunisia, December 16-18, 2008.

[2] A. Merchant and B. Sengupta, "Assignment of cells to switches in PCS networks," IEEE/ACM Transactions on Networking 3 (5), 521–526 October, 1995.s

[3] A. Mukherjee, S. Bandyopadhyay, and D. Saha, "Location Management and Routing in Mobile Wireless Networks, Mobile communication Series, Artech House INC., ISBN:1-58053-355-8, 2003.

[4] A.Quintero and S. Pierre, "Evolutionary Approach to optimize the assignment of Cells to Switches in Personal Communication Networks," Wireless Communications Vol.26 (9) pp. 927-938, June, 2003.

[5] C. Hedible and S. Pierre,"A Genetic Algorithm for Assigning Cells to Switches in Personal Communication Networks," IEEE Canadian Review 44, pp. 21–24, 2003.

[6] I. Demirkol, C. soy, M. Caglayun, and H. Delic, "Location

Area Planning in Cellular Networks Using Simulated Annealing," Proc. INFOCOM, pp. 566-573, March 2001.

[7] 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, Nagoya, Japan, pp. 39-43,1995.

[8] R. Fournier and S. Pierre,"Assigning Cell to Switches in Mobile Networks using an Ant Colony Optimization heuristics," Computer Communications28, pp. 65-73, 2005.

[9] J. Kennedy, R. Eberhart, "A Discrete Binary Version of the Particle Swarm Algorithm," IEEE international conference on Systems, Man, and Cybernetics, pp. 4104 – 4108, 1997.

[10] Y. Chun-man, G. Bao-long, W. Xian-xiang, "Empirical Study of the Inertia Weight Particle Swarm Optimization with Constraint Factor," International Journal of Soft Computing And Software Engineering (JSCSE) ISSN: 2251-7545 Vol.2, No2, 2012.

[11] K. Udgata1, U. Anuradha, G. Pavan Kumar," Assignment of Cells to Switches in a Cellular Mobile Environment using Swarm Intelligence," IEEE, International Conference on Information Technology (ICIT), pp. 189-194, 2008.

International Journal of
Intelligent Systems and
Applications in Engineering
ISSN:2147-6799 www.atscience.org/IJISAE
Advanced Technology and Science
Brief Report

About a discussion ‘‘Development a new mutation operator to solve the Traveling Salesman Problem by aid of genetic algorithms’’, by Murat Albayrak and Novruz Allahverdi, 2011. Expert System with Applications, 38; 3, pp. 1313–1320.

Novruz Allahverdi[*]

Received 04th June 2014, Accepted 01th April 2015

Abstract: In the Short Communication published in “Expert Systems with Application” in volume 41 2014, (Comments on "Albayrak, M., & Allahverdi N. (2011). Development a new mutation operator to solve the Traveling Salesman Problem by aid of Genetic Algorithms. Expert Systems with Applications, 38(3), 1313-1320" [1]: A Proposal of Good Practice; E. Osaba, E. Onieva, F. Diaz, R. Carballedo, Volume: 41, Issue: 4, Pages: 1530-1531, Part: 1, March 2014) [4] the Osoba E. et al have discussed our method to solve the Traveling Salesman Problem pointing that we use our developed new algorithm to compare different versions of a classical genetic algorithm, each of one with a different mutation operator and they write that this can generate some controversy. Here we shortly analyze the comment of Osaba E. et al. to show that our comparing method has a chance of existence.

Keywords: *Genetic algorithms, Traveling Salesman Problem, algorithm Greedy Sub Tour Mutation (GSTM).*

1. Analysis:

We can classify the proposals Osaba E. and others four class:

(1) Comparison and evaluation of the Greedy and Normal mutation methods together are not correct (it is wrong).
As stated in our article "Development a new mutation operator to solve the Traveling Salesman Problem by aid of Genetic Algorithms" [1] our new mutation algorithm Greedy Sub Tour Mutation (GSTM) has a hybrid structure. GSTM operator acts as a greedy, at the same time include the operators of Simple Inversion Mutation (SIM) and Scramble Mutation (SCM). Also if you look at the values of PRC = 0.5, PCP = 0.8, as used in our analysis it can be seen that the probability of using GSTM classical operators is larger. In this case we can say that the comparison of operators GSTM greedy and classic is applied properly.

(2) Compare with Non-Sequential 4-Change that is described in literature (Freisleben & Merz (1996) [2].
It is not logical to compare the performance GSTM, and Non-Sequential 4-Change (Double Bridge Kick Move) techniques described in Freisleben & Merz (1996), as recommended by the Osaba & et. [4]. Non-Sequential 4-Change operator is not used singly as a mutation operator. This operator is also used to perturbations and then created a new generation to be having a local search by Lin-Kernighan method, which is a sort of heuristic improvement. So the quality of the resulting new generation is determined by the Lin-Kernighan method. GSTM which we have described in the article [1] acts as greedy and also includes a natural hybrid mutation operation and is not a local search method. Therefore, to compare our method with the mutation method developed in this article is not proper.

(3) It is confirmed that all greedy methods are used together (NN + DPX). So which of these methods have a success is not clear.
All of Genetic Algorithms in the analysis Table 1 and Table 2 presented in our article the primary population (Nearest Neighbor) and crossover method (DPX) were chosen the same way. The only option is different: mutation methods. In this way, the impact of mutation methods in GA was observed. From Tables 1 and 2, it is clear that the reason for getting the best value error (%) and average error (%) obtained in the GSTM arises from the operator developed by us. Even if you use a standard crossover method and random initial population in GA worked with a level of time equal intervals as shown in Table 2, is not difficult to see that the use of other methods of mutation gives a higher value of the error and the difference between GSTM and the other methods will increase.

(4) Generate primary population by randomly and make OX crossover and test again…
We can analyze the methods proposed in the Osaba and coauthor’s papers [3] and [5] in this way:
In this regard, we can say that performance some of naturel crossover operators (OX+CX+PMX) in GSTM was investigated and very good results comparing with other mutation operators were submitted as graphics. This case proves that GSTM performance is not depend on greedy crossover methods and it (GSTM) can demonstrate high performance with naturel crossover methods.

2. Conclusion

[*] *Selcuk University, Faculty of Technology, Department of Computer Engineering, Konya, Turkey.*
email:noval@selcuk.edu.tr

So, we believe that our comparing method has a chance of existence. The proposal of comparing our and Osaba E. and coauthors algorithms is interesting, but they have to test it themselves, because their paper was published after our paper.

3. Comment

When this article was prepared for printing Osaba and coauthors published some new works. In these studies, they give the reference to the algorithm developed by us [1] and they say: "Some other operators of this type can be … the greedy sub-tour mutation" [6] and "Normally, these operators are heuristic, and they are applied to a particular problem, in which they get a great performance" [7].

References

[1] [1] Albayrak, M., & Allahverdi N. (2011). Development a new mutation operator to solve the Traveling Salesman Problem by aid of genetic algorithms. Expert Systems with Applications, 38(3), 1313–1320.

[2] [2] Freisleben, B., & Merz, P. (1996). A genetic local search algorithm for solving symmetric and asymmetric traveling salesman problems. In Proceedings of IEEE international conference on evolutionary computation (pp. 616–621), IEEE.

[3] [3] Osaba, E., Carballedo, R., Diaz, F., & Perallos, A. (2013). Analysis of the suitability of using blind crossover operators in genetic algorithms for solving routing problems. In Proceedings of the 8th International Symposium on Applied Computational Intelligence and Informatics (pp. 17–23). IEEE.

[4] [4] Osaba E., Onieva E., Diaz F., Carballedo R. Comments on "Albayrak, M., & Allahverdi N. (2011). Development a new mutation operator to solve the Traveling Salesman Problem by aid of Genetic Algorithms. Expert Systems with Applications, 38(3), 1313-1320": A Proposal of Good Practice; Volume: 41, Issue: 4, Pages: 1530-1531, Part: 1, March 2014)

[5] [5] Osaba, E., & Díaz, F. (2012). Comparison of a memetic algorithm and a tabu search algorithm for the traveling salesman problem. In Federated conference on computer science and information systems (FedCSIS), pp. 131–136, IEEE.

[6] [6] Osaba E., Carballedo R., Diaz F., Onieva E., Lopez P., Perallos A. On the influence of using initialization functions on genetic algorithms solving combinatorial optimization problems: a first study on the TSP. 2014 IEEE Conference on Evolvıng and Adaptıve Intellıgent Systems (EAIS 2014), Linz, Austria; 06/2014.

[7] [7] Osaba E., Carballedo R., Diaz F., Onieva E., I. de la Iglesia, Perallos A. (2014), Crossover vs. Mutation: A Comparative Analysis of the Evolutionary Strategy of Genetic Algorithms Applied to Combinatorial Optimization Problems. The Scientific World Journal, August, 2014, 22 p.

International Journal of
Intelligent Systems and
Applications in Engineering
ISSN:2147-6799 www.atscience.org/IJISAE

Advanced Technology and Science

Original Research Paper

New Approach in E-mail Based Text Steganography

Kemal Tutuncu *[1], Abdikarim Abi Hassan

Accepted 26th May 2015

Abstract: In this study combination of lossless compression techniques and Vigenere cipher was used in the e-mail based text steganography. It makes use of email addresses to be the keys to embed/to extract the secret message into/from the email text (cover text). After selecting the cover text that has highest repetition pattern regarding to the secret message the distance matrix was formed. The members of distance matrix were compressed by following lossless compression algorithms as in written sequence; Run Length Encoding (RLE) + Burrows Wheeler Transform (BWT) + Move to Front (MTF) + Run Length Encoding (RLE) + Arithmetic Encoding (ARI). Later on Latin Square was used to form stego key 1and then Vigenere cipher was used to increase complexity of extracting stego key 1. Final step was to choose e-mail addresses by using stego key 1 (K1) and stego key 2 (K2) to embed secret message into forward e-mail platform. The experimental results showed that proposed method has reasonable performance in terms of capacity and also higher security in terms of complexity.

Keywords: *Text steganography, Latin square, Vigenere chipher, Stego key, BWT, MTF, RLE, ARI*

1. Introduction

The case of safety and security of information and especially secret communications has led to the introduction of several methods for secret message. Among these methods steganography is a rather new method. Steganography, the art of information hiding, has been used around for thousands of years, with the earliest examples coming from as early as 440 B.C.

Steganography is the traditional method used in information hiding. The use of systems like cryptography it makes the output of these methods very conspicuous. The main reason behind this is that the systems make messages unreadable to anyone except the intended recipient and does not hid its existence. The design of steganographic approaches is aimed at hiding the message or (data) and ensuring that its existence is hidden from observers.

Modern steganography can be applied to text, images, audio and video as shown in Fig.1. However, text steganography, has received less interest recent years, primarily due to the difficulty in finding redundant bits in text files and the lower capacity to hide information than the other mediums. This should not be the issue, as text Steganography has many advantages over the other mediums which make it model for effective Steganography.

One advantage of text Steganography over images and audio is that while they are both susceptible to compression due to their use of redundant data, this is not an issue with text Steganography as even though text contains redundancy, it cannot be removed or compressed. In addition the advantage to prefer text Steganography over images and audio is its smaller memory occupation and simpler communication. Text is also still one of the major forms of communication in the world, both in digital and printed form, and there are not many people who do not have access to text.

Text Steganography can be classified into three main groups: Structural, random and statistical generational, and finally linguistic. The main characteristics of structural text Steganography is that it modifies the physical form of the text, for instance through the appending white spaces and linespaces. On the other hand, random and statistical generation entails providing the cover text. This can be in a random manner or depending on specified input. While linguistic Steganography uses contents of natural language, for example verbs, nouns, adjectives and so on. Linguistic Steganography are classified into two: syntactic and semantic Syntactic text Steganography deals with changing the format of the text without considerable alteration to the meaning or tone.

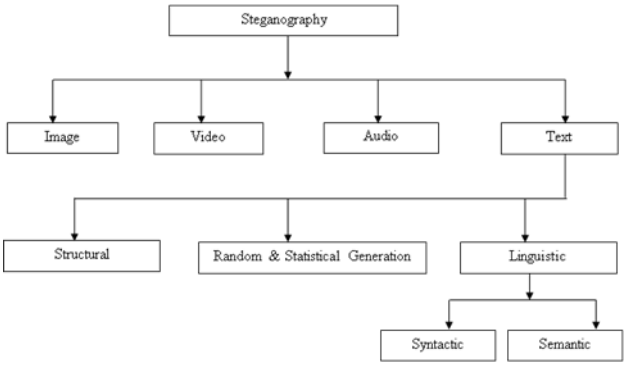

Figure 1. Categories of steganography

In this study, an email based data hiding method which makes use of combinatorial lossless compression to increase the hiding capacity and Vigenere cipher to increase the complexity was proposed. Run Length Encoding (RLE) + Burrows Wheeler Transform (BWT) + Move to Forward (MTF) + Run Length Encoding + Arithmetic Encoding (ARI) algorithms were used to have better compression ratio. Additionally Vigenere cipher was used to increase complexity of the system. The secret data was embedded into the email ids of forward mail platform. The most appropriate cover text is chosen from the text base by using distance matrix. After the stego cover determination we make use of distance matrix to find out email ids of a forward mail platform. Then e-mails were chosen from the previously arranged email address list. This email address list is used as a global stego key that is shared between both the sender and the receiver in advance. Capacity metric of the proposed method was evaluated by using obtained e-mail addresses.

The rest of the paper is structured as follows. Section II provides the related works. Section III and Section IV describe the proposed method and the experimental result. Finally, the

[1] *Faculty of Technology, Selcuk University Campus,Konya,Turkey*
** Corresponding Author: Email: ktutuncu@selcuk.edu.tr*

conclusion is drawn in Section V.

2. Related Works

In [1], [2], Satir and Isik used email to be the cover of the secret data. Sending same e-mail to many recipients were the key idea of their studies. There may not be a proper method to validate whether the email address of each recipient is valid but sending an email to the respective address. In those papers [1], [2], email addresses are used to hide the stego keys, regardless they are valid addresses or not [3]. In [3] the authors proposed improvement of the study [2], such as the randomity of the generated email addresses. They claimed that random generated email addresses may increase the security level. In [4], the authors proposed combination of BWT + MTF + LZW coding algorithm to increase the hiding capacity and adding some random characters just before the '@' symbol of email ids to increase the randomness.

In a chat application, Wang et al. [5] employ emoticons (emotional icons) to hide a secret message. This is intended to improve the performance of other chat-based steganography. The proposed method is developed based on the assumption that the use of emoticons in chatting is high, where many recent applications using it. It is claimed that this method is able to raise up the capacity of the secret data to be embedded as well as to provide an easy to use application. This method requires both sender and receiver have an exactly same emoticon table containing some emoticon sets. In this case, the order of the emoticon affects the meaning of the secret message. Therefore, there must be a synchronization process before the communication begins. Practically, this method requires the users to input many emoticons in order to send the secret data. This may have an effect on the user convenience [4]. Wayner [6,7] discussed an important method using mimic functions. It applies the inverse of Huffman Code having employed the randomly distributed bits of input stream on itself. Other researches have been carried out which employ varied communication medium, such as in [8], [9]. Hiding the secret in the text is also investigated by Desoky [10]. He proposes Listega, a method to hide data which exploits the popularity of list of items, such as song, food, drink and car. Another text steganography given by Maher [11] which is popular as TEXTO is built to transform uuencoded or PGP ASCII armored ASCII data into English language sentences. It converts the secret data into English words. To extend the work of [12], another important method is synonyms-based approach [13-15]. Unlike [11], the method uses legitimate words and sentences having appropriate preciseness. Thus, the visual attack will have wee significance on these types of methods.

3. Proposed Method

3.1. Embedding Phase

Let, S: secret message T: A text base of cover texts in which the secret message will be embedded. K1: set of email addresses having four characters before '@' symbol. It is shared between the sender and the receiver. A: set of the second parts of email addresses such as outlook.com, gmail.com, etc.

Step1. Construct difference matrix D in order to select most suitable or relevant text from the T. the D is calculated by having difference of index of last matched symbol of S with current index of match symbol in T iteratively except for the first symbol of S as in case of first symbol, the last symbol index is 0 as in [1,2,3,4]. **Step2.** Calculate vectors R and E using following equations: R=D mod 26 E=D/26

Step3. Estimate the maximum dual pattern repetition in R and store in a column matrix P. Now, select the largest row of P and denote it as Pmax. The corresponding rows of R and E vectors are also selected and put in R* and E* vectors. The text from T corresponding to Pmax is also chosen and put in T* as it is the most suitable text.

Step4. Apply Run Length Encoding (RLE) + Burrows Wheeler Transform (BWT) + Move to Forward (MTF) + Run Length Encoding + Arithmetic Encoding (AE) in the written order for compressing elements of R*

Step5. Represent each element of R in binary form and concatenate them in order to obtain bit stream.

Step6. The bit stream is partitioned into groups of 12 bits, in each group, the first 9 bits are called G1, next 3 bits are called G2. The quotient and remainder of decimal representation of G1 with respect to 26 are known as x and y respectively and decimal representation of G2 is known as z.

Step7. Choose two characters as K1 by employing Latin square on x and y. Select extensions of email addresses using z from A.

Step8. Use Vigenere cipher to alter two characters obtained from Latin Square. Thus these two new characters will be used to choose e-mail address. Choose email addresses that starts with modified K1 before @ and ends with z after @.

Step9. Modify resultant email addresses in order to complete construction of K2 set by using E*. Combine modified K1 and K2 (obtained from z) to form e-mail addresses.

3.2. Extracting Phase

Before starting extraction phase both the sender and receiver must have email address list, three bits corresponding of email extension, Vigenere cipher key, RLE, MFT, BWT and ARI tables in advance.

Step1. Get the stego-cover. Extract numeric elements of K2 before '@' symbol to construct the vector E. If there is not any numeric element then E will be 0.

Step2. Extract first two elements from K2 to obtain Vigenere values of x and y by employing Vigenere Cipher.

Step 3. Extract the values of x and y by employing Latin Square Also extract email address extension to obtain z.

Step 4. Calculate G1 and G2 for each group of 12 bits by using the following equations: G1 = (x • 26+y) G2 = (z)

Step4. Concatenate G1 and G2 in same order to obtain compressed bit stream.

Step5. Decompress the bit stream using Arithmetic Encoding (AE) + Run Length Encoding (RLE) + Move to Forward (MTF) + Burrows Wheeler Transform (BWT) + Run Length decoding to obtain original R*

Step7. Estimate original difference D using R* and E as follows: D = R + (26 • E)

Step8. By using elements of D, extract the elements of S through T*, in the stego cover.

4. Results

The proposed method was prepared by programming in MATLAB. The computer used for running the program was Intel Pentium7, with 8 GB RAM. Later on same computer was used to analyse performance of proposed method or program. The performances of the text steganography methods can be analysed by capacity ratio or complexity. In this study both parameters are investigated. In order to calculate the capacity ratio secret data whose length is 196 characters is used. The secret message is as follows:

'behind using a cover text is to hide the presence of secret

messages the presence of embedded messages in the resulting stego text cannot be easily discovered by anyone except the intended recipient'. Cover data set was ranging from 700 characters to 4000 characters. 35 cover data set was used. It has been seen that the success of the embedding is directly related with length of the secret message and cover text. The smaller size the cover text the less successful the embedding is. Two factors are important here; Firstly the length of cover to possibly include all the characters of the secret message due to its length and secondly the match between the content of the cover message with the frequencies off English alphabet. If a cover text is formed according to the frequency of English alphabet it will cause better capacity ratio for text steganography.

Table 1. Applying text message (200 char) to different cover text

Cover	Cover Length (char)	Result	Number of e-mail addresses
CT1	300	Fail	-
CT2	800	Fail	-
CT3	1500	Fail	-
CT4	3993	Success	1957

As can be seen from Table 1. CT1, CT2 and CT3 cover texts resulted in not hiding secret message. No e-mail address is produced to hide secret text. CT4 successfully hided the secret text. Here the key point is how much repetition occurs when we created distance matrix. If we have much more repeated number in distance matrix then the compression ratio of the combined lossless compression techniques will increase. This will cause increase in capacity ratio, thus decrease number of e-mail addresses. From this point of view it is quite important to choose cover text to increase the capacity ratio in text steganography. Briefly explaining not only the length of the cover data determines the successfulness of the embedding process, but also variation of the characters. Longer the cover data actually increases the possibility of providing more varied characters. Therefore, successfulness of the embedding process depends on the appropriate selection of the cover data [3].

Complexity is one of the important aspects of the steganography. It is directly related with extraction of hidden information by an adversary in an easy way or not. In this study compression algorithms and two combinatory based coding (Latin square and Vigenere cipher) were used to increase the complexity. Complexity analysis of the proposed system will be explained. Before explanation one must keep in mind that global key and the key of Vigenere cipher were shared between sender and receiver in advance. Thus, if any adversary would like to break the system he will not have these two keys. The step that an adversary will follow to break the system as follows:

1. Adversary must extract compressed R*. In order to implement this action he/she has to obtain G1from x, y and G2 from z. At his point x and y can only be obtained by using the Vigenere cipher that consists of two characters. Finding correct combination of these two characters will require 26*26 combinations. Finding correct combination will let adversary to find the Latin Square corresponding of x and y. Afterwards x and y can be obtained from Latin Square. The z can be obtained by the extension of e-mail address. Adversary has to calculate 8 combination per each e-mail. Thus, if the number of e-mail address is 8 then 8^N combination must be calculated. Later on each z is expressed in binary form and then combined with x and y to obtain G1 and G2. In order to find out correct bit stream R*, 8^N*26^2 combination must be formed.

2. R is obtained by solving R*. As it is known that D = R + (26 • E) and each e-mail address is checked to see whether it has number before @ sign or not. In case of including numbers before @ sign these numbers are used for obtaining members of E. These numbers can belong to e-mail address naturally or added to e-mail address to hide members of E during embedding phase. If we express number of numbers in an e-mail address as m then the adversary will implement 2^m combination to extract the members of E.

3. Obtained each R and E must be tested to each other to see the correct hidden message.

Thus complexity of the proposed system is

$Comp\_proposed\_system = 26^2\ 8^N \times \prod_1^N 2^m$

Complexity of the system is calculated without regarding the compression algorithms that are used to have combination of RLE + BWT + MTF + RLE + AE. In the study of [2] the authors didn't make use of complexity of obtaining decompressed text. In order to make comparison with this study only the complexities contribution of k1, k2 and Vigenere cipher were used. In [2] complexity was followed as $8^N \times \prod_1^N 2^m$. Thus proposed method increased the complexity as the amount of 26^2 .

5. Conclusion

Proposed method is extension of text-based steganography that makes use of email environment to hide the secret text. Capacity ratio and complexity of the proposed method are improvements of the previous e-mail environment based text steganography methods [1-4] in terms of complexity. Reasonable capacity was obtained. By using combination of Run Length Encoding (RLE) + Burrows Wheeler Transform (BWT) + Move to Front (MTF) + Run Length Encoding (RLE) + Arithmetic Encoding (ARI) algorithms the size of the secret message was reduced. Vigenere cipher added extra complexity or security to the system for obtaining stego key (K1). The study also showed that the length and the content (variation) of the cover text directly related with the success of the embedding process. Short cover text with not well distributed alphabet content (doesn't match with the frequency of the English alphabet) will result failure in embedding process.

Other combination of lossless compression techniques can be tried for increasing the capacity ratio of the email environment based steganography. Additionally some other techniques that put randomness to the email addresses can be used to increase the complexity of the related system.

References

[1] E. Satir and H. Isik, A compression-based text steganography method, The Journal of Systems and Software Science Direct, vol. 85, issue 10, pp. 2385-2394, 2012.

[2] E. Satir and H. Isik, A Huffman Compression based Text Steganography Method, Multimedia Tools Appl, September 2012.

[3] Tohari Ahmad, Melvin S. Z. Marbun, Hudan Studiawan, Waskitho Wibisono, and Royyana M. Ijtihadie, A Novel Random Email-Based Steganography International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 4, No. 2, April 2014

[4] Rajeev Kumar , Satish Chand , and Samayveer Singh, An

Email based high capacity text steganography scheme using combinatorial compression, 25th International Conference-Confluence The Next Generation Information Technology Summit (Confluence), 25-26 September 2014, India

[5] Z. Wang, T. Kieu, C. Chang, and M. Li, Emoticon-based text steganography, in Proc. 2009 Asia-Pacific Conference on Computational, Wuhan, China, 2009.

[6] P. Wayner, Mimic Functions, Cryptologia vol. 16(3), pp. 193-214, 1992.

[7] P. Wayner, Disappearing Cryptography, AP Professional, Chestnut Hill, MA (1996).

[8] L. Por, K. Wong, and K. Chee, UniSpaCh: a text based data hiding method using unicode space characters, Journal of Systems and Software, vol. 85, issue 5, pp. 1075-1082, 2012.

[9] M. Topkara, U. Topkara, and M. Atallah, Information hiding through errors: a confusing approach, in Proc. SPIE International Conference on Security,Steganography, and Watermarking of Multimedia Contents, San Jose, CA, 2007

[10] A. Desoky, Listega: list-based steganography methodology, International Journal of Information Security, vol. 8, no. 4, pp. 247-261, 2009.

[11] K. Maher, TEXTO. 1995. ftp://ftp.funet.fi/pub/crypt/steganography/texto.tar.gz.

[12] K. Winstein, "Lexical steganography through adaptive modulation of the word choice hash", Secondary education at the Illinois Mathematics and Science Academy, January 1999.

[13] H. Nakagawa, K. Sanpei, T. Matsumoto, T. Kashiwagi, S. Kawaguchi, K. Makino and I. Murase, "Meaning Preserving Information Hiding _Japanese text Case," IPSJ Journal, Vol.42, No.9, pp. 2339 - 2350, 2001. (In Japanese)

[14] B. Murphy and C. Vogel, The syntax of concealment: reliable methods for plain text information hiding, In Proceedings of the SPIE Conference on Security, Steganography, and Watermarking of Multimedia Contents, San Jose, CA, vol. 65(05), 2007.

[15] A. Desoky, Listega: List-Based Steganography Methodology, International Journal of Information Security, Springer-Verlag, vol. 8, pp. 247-261, April 2009.

