



Penev, Kalin and Peytchev, Evtim and Mihaylova, Lyudmila and Fidanova, Stefka and Roeva, Olympia and Iliev, Georgi and Vasileva, Vesela and Lazarova, Mariya and Simeonov, Ivaylo and Yalamova, Gergana and Tomov, Yavor and Shindarov, Miroslav and Marinov, Pencho and Slavov, Tsonyo and Gamati, Emadeddin and Li, Yueyue and Yasser, Alroqi and Zhang, Qiming and Ali, Arshad and Xydeas, Costas and Umer, Tariq and Ahmad, Hassan and Honary, Bahram and Honary, Mahsa and Essafi, Daniel. (2012). Optimisation of Mobile Communication Networks - OMCO NET. Southampton Solent University, UK. ISBN 978-0-9563140-4-8

Downloaded from <http://ssudl.solent.ac.uk/2214/>

Usage Guidelines

Please refer to usage guidelines at <http://ssudl.solent.ac.uk/policies.html> or alternatively contact ir.admin@solent.ac.uk.

Mini-conference

**Optimisation
of Mobile Communications
Networks**

28-30 June 2012

Southampton

Proceedings

Mini Conference – Optimisation of Mobile Communication Networks

28-30 June 2012

Technology School, Southampton Solent University

Editor

Kalin Penev

British Library Cataloguing in Publications Data

Penev, K.

Optimisation of Mobile Communication Networks

ISBN 978-0-9563140-4-8

All rights reserved. This publication may not be translated or copied in whole or in part without the prior permission in writing of the Editor except for fair dealing for the purpose of research or review, as permitted under the Copyright, Design and Patents Act 1988. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology known or hereafter developed is strictly forbidden. Enquiries concerning reproduction outside those terms should be sent to the Editor.

The use of registered names, trademarks etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any responsibility or liability for any errors or omissions that may be made.

Typesetting and electronic files prepared by Vesela Vasileva.

Printed and bound by CLOC Ltd., UK

Chair

Kalin Penev, Southampton Solent University, UK

Organising Committee

John Rees, Southampton Solent University, UK

Kevin Walsh, Southampton Solent University, UK

Vesela Vasileva, Southampton Solent University, UK

Program Committee

Stefka Fidanova, Bulgarian Academy of Sciences, Bulgaria

Alexander Gegov, University of Portsmouth, UK

Georgi Iliev, Technical University of Sofia, Bulgaria

Lyudmila Mihaylova, University of Lancaster, UK

Evelina Pencheva, Technical University of Sofia, Bulgaria

Evtim Peytchev, Nottingham Trent University, UK

Olympia Roeva, Bulgarian Academy of Sciences, Bulgaria

Acknowledgements

Optimisation of Mobile Communications Networks – Mini Conference and preparation of these proceedings are supported by Southampton Solent University, Research and Enterprise Fund, Grants 516/17062011, 623/19102011.

Content

Special Session

Optimisation Methods – Novel Aspects

Dr Kalin Penev, Southampton Solent University7

Keynote: Algorithms Applied to Global Optimisation – Visual Evaluation

Kalin Penev8

Free Search and Particle Swarm Optimisation Applied to Non-constrained Test

Vesela Vasileva, Kalin Penev20

Differential Evolution with Enhanced Abilities for Adaptation Applied to Heterogeneous Numerical Optimization

Mariya Lazarova, Kalin Penev28

Self-Adapting, Multi-Parent Recombination Evolution Strategy Algorithm

Ivaylo Simeonov, Kalin Penev 34

Real Coded Genetic Algorithm with Enhanced Abilities for Adaptation Applied to Optimisation of MIMO Systems

Gergana Yalamova, Georgi Iliev, Kalin Penev41

Overview of Routing Algorithms for MCP with Single or Mixed Metric

Yavor Tomov 50

Mono-objective Algorithm for Wireless Sensor Layout

Stefka Fidanova, Miroslav Shindarov, Pencho Marinov57

Application of Simulated Annealing and Threshold Accepting for a Parameter Tuning Problem

Olympia Roeva, Tsonyo Slavov64

Special Session

Wireless Networks – Collaborative Information and Control Systems

Dr Evtim Peytchev, Nottingham Trent University72

Collaboration in Wireless Networks – Towards New Paradigm in Building Intelligent Traffic Systems

Evtim Peytchev73

Sensing Traffic Conditions in VANET for Building New Generation in Wireless Intelligent Transportation Information Systems

Emadeddin Gamati, Evtim Peytchev74

Traffic Message Delivery Broadcast Protocol in Vehicular Ad Hoc Networks

Yueyue Li, Evtim Peytchev83

Mobile Network Architecture for Application Support in Remote Location

Yasser Alroqi, Evtim Peytchev91

Traffic Simulation Models Review

Qiming Zhang, Evtim Peytchev96

Special Session

Localisation and Positioning in Sensor Networks

Dr Lyudmila Mihaylova, University of Lancaster103

Keynote: Advances in Bayesian Methods for Tracking and Decision Making

Lyudmila Mihaylova104

Sensor Position Management based on Interpolation in Wireless Sensor Networks

Arshad Ali, Costas Xydeas, Lyudmila Mihaylova108

Stochastic Model for Improving Connectivity under Heterogeneous Traffic Flow for VANET

Tariq Umer, Hassan Ahmad, Bahram Honary118

Enhanced Positioning through Inference of Mobile Terminal Context

Mahsa Honary, Daniel Essafi, Lyudmila Mihaylova126

Special Session

Optimisation Methods – Novel Aspects

Keynote: Algorithms Applied to Global Optimisation – Visual Evaluation
Dr Kalin Penev, Southampton Solent University

Algorithms Applied to Global Optimisation – Visual Evaluation

Kalin Penev
Southampton Solent University, Southampton, UK
Kalin.Penev@solent.ac.uk

Abstract: Evaluation and assessment of various search and optimisation algorithms is subject of large research efforts. Particular interest of this study is global optimisation and presented approach is based on observation and visual evaluation of Real-Coded Genetic Algorithm, Particle Swarm Optimisation, Differential Evolution and Free Search, which are briefly described and used for experiments. 3D graphical views, generated by visualisation tool VOTASA, illustrate essential aspects of global search process such as divergence, convergence, dependence on initialisation and utilisation of accidental events. Discussion on potential benefits of visual analysis, supported with numerical results, which could be used for comparative assessment of other methods and directions for further research conclude presented study.

Keywords: Search Process Visualisation, Real-Coded Genetic Algorithm, Particle Swarm Optimisation, Differential Evolution, Free Search, Numerical Optimization.

1. Introduction

Global optimization refers to finding optimal solution of a given non-convex objective function [6][14]. Real world tasks are often global and need reliable methods to cope with. For this purpose various search methods such as Genetic Algorithm [7], Particle Swarm Optimization (PSO) [2][3], Differential Evolution [13] and Free Search (FS) [9][10] can be used. However majority of search and optimization methods face difficulties when dealing with global optimization problems. The main reasons of their failure are: trap in local sub-optimal solution, inability to escape from trapping, inability to abstract appropriate knowledge or use it effectively (if available).

Observation of optimisation process and visual analysis significantly help to identify dependence on initialisation of some methods, abilities to diverge across the whole search space, abilities to converge to optimal solution, abilities to use accidental events and to abstract from them knowledge, which could facilitate search process. This study uses for experiments Real-Coded Genetic Algorithm, Particle Swam Optimisation, Differential Evolution and Free Search.

1.1. Genetic Algorithm

Genetic Algorithms are computational models inspired by the concept about natural selection and evolution of biological species. Natural evolution can be considered as a kind of search process [7]. Therefore this concept is recognised as valuable in the domain of heuristics optimisation and search methods. A computational implementation and application of Genetic Algorithms are published in the literature [7]. Genetic algorithms are different from other optimisation and search processes in several ways: (1) GAs work with a coding of the parameter set, not the parameters themselves; (2) GAs search from a population of points, not from a single point; (3) GAs use payoff (objective function) information, not derivatives or other auxiliary knowledge; (4) GAs use probabilistic transition rules, not deterministic rules [5]. A GAs major event is modification. It involves selection of parents, recombination between them, mutation and evaluation. For this study a Blend crossover modification strategy called BLX- α [4] is selected.

1.2. Particle Swarm Optimisation

PSO can be classified as a population-based, evolutionary computational paradigm [2]. It has been compared to Genetic Algorithms [1][3] for efficiently finding optimal or near-optimal solutions in large search spaces. PSO is different from other evolutionary computational methods. It attempts to model a social behaviour of a group of individuals [2][12]. In PSO each particle is defined as a potential solution to a problem in multi-dimensional space. One of the advantages of PSO is flexible tuning of few parameters. One version, with slight variations, works well in a wide variety of applications. A variable called inertia factor influences PSO positively. Large inertia factor facilitates global exploration and searching new areas, while small inertia factor tends to facilitate local exploration and fine-tunes the current search area [5].

1.3. Differential Evolution

Differential Evolution can be described as real-value method for optimising non-linear and non-differentiable functions within continuous space [11][13]. It starts with stochastic selection of an initial set of solutions called design vectors. The value of an objective function, which corresponds to each individual of the population, is a measure of that individual's fitness as an optimum. Then, guided by the principle of survival of the fittest, the initial population of vectors is transformed, generation-by-generation, into a solution vector. DE selects for manipulation target, donor and differential vectors. Therefore the minimal number of vectors in one population has to be more than four. For modification strategies, which use four differential vectors the minimal population size is seven. The current target and corresponding new trial vector (individual) in each generation are subject of competitions to determine the composition of the next generation. The new trial vector is generated in several steps as follows: (1) selection of a randomly chosen donor vector from the population different from the current target vector; (2) selection of other (two or four) randomly chosen vectors (so called differential vectors), different from the donor, different from the current target vector and different from each other; (3) calculation of a difference between differential vectors and scaling it by multiplication with a constant called differential factor; (4) adding the difference to the donor vector, which produces a new vector; (5) crossover between the current target vector and the new vector so that the trial vector inherits parameters from both of them. If the trial vector is better than the current target vector, then the trial vector replaces the target vector in the next generation. In all, three factors control evolution under DE: the population size; the scaling weight (differential factor) applied to the random vectors differential.

1.4. Free Search

Free Search is real-value adaptive heuristic method. The search process is organised in exploration walks, which differs from classical iterations [9][10]. It starts with initialisation. The algorithm requires definition of the search space boundaries [$Xmin_i$ and $Xmax_i$], population size m , limit for the number of explorations G , limit for the number of steps for exploration T , minimal and maximal values for the frame of a neighbourhood space [$Rmin$, $Rmax$]. The maximal neighbour space $Rmax$ guarantees coverage of the whole search space by one individual. The minimal neighbour space $Rmin$ guarantees desired granularity of the coverage by one individual. $Rmin$ and $Rmax$ are absolute values. An appropriate definition of these values supports good performance across a variety of problems without additional external adjustment. A prior determination of the neighbour space and preliminary adjustment of the algorithm for a particular problem based on preceding knowledge can lead to slightly better performance on that problem but aggravates the performance on other problems which concurs with the existing general assessment of the performance of optimisation algorithms [10].

FS requires definition of an initialisation strategy. Acceptable initialisation strategies are:

$$\text{- random values: } x_{0ji} = Xmin_i + (Xmax_i - Xmin_i) * random_{ji}(0,1), \quad (1)$$

$$\text{- certain values: } x_{0ji} = a_{ji}, \quad a_{ji} \in [Xmin_i, Xmax_i], \quad (2)$$

$$\text{- one location: } x_{0ji} = c_i, \quad c_i \in [Xmin_i, Xmax_i], \quad (3)$$

where $random(0,1)$ is a random value between 0 and 1, a_{ji} and c_i are constants.

The ability to operate with all these strategies also supports good performance across a variety of problems without constant re-tuning of internal operator parameters. For multi-start optimisation FS allows variation of the initialisation strategies. Upon initialisation each individual takes an exploratory walk. It generates coordinates of a new location x_{tji} as:

$$x_{tji} = x_{0ji} - \Delta x_{tji} + 2 * \Delta x_{tji} * random_{tji}(0,1). \quad (4)$$

Modification strategy used in the algorithm is:

$$\Delta x_{tji} = R_{ji} * (Xmax_i - Xmin_i) * random_{tji}(0,1), \quad (5)$$

where $i = 1$ for a one-dimensional step (i indicates one dimension); $i = 1, \dots, n$ for a multi-dimensional step. t is the current step $t = 1, \dots, T$. T is the step limit per walk. R_{ji} indicates the size of the idealised frame of the neighbourhood space for individual j within the dimension i . $random_{tji}(0,1)$ generates random values between 0 and 1. Δx_{tji} indicates the actual size of the neighbourhood space for a particular problem for step t of individual j within dimension i . During the exploration an individual with a neighbourhood space, which exceeds search space boundaries, can perform global exploration whereas another individual with small neighbour space can make precise steps around one location.

Modification strategy is independent from the current or the best achievements. The exploration performs heuristic trials based on stochastic divergence from one location. The concrete value of the neighbourhood space for a particular exploration defines the extent of uncertainty of the chosen individual. The walk is followed by an individual assessment of the explored locations. The best location is marked with pheromone. The pheromone indicates the locations quality and may be described as result or cognition from previous activities. The assessment, during the exploration, is defined as follows:

$$f_{tj} = f(x_{tji}), \quad f_j = \max(f_{tj}), \quad (6)$$

where f_{tj} is the value of the objective function achieved from animal j for step t , f_j is the quality of the location marked with pheromone from an individual after one exploration.

The pheromone generation is generalised for the whole population:

$$P_j = f_j / \max(f_j), \quad (7)$$

where $\max(f_j)$ is the best achieved value from the population for the exploration.

This is a normalisation of the explored problem to an idealised qualitative (or perhaps cognitive) space, in which the algorithm operates. This idealised space uses for a model an idealised space of notions in thought of biological systems, in which they generate decisions. The normalisation of any particular search space to one idealised space supports automation and successful performance across variety of problems without additional external adjustments.

Then a generation and a refining of the sensibility follow. The sensibility generation is:

$$S_j = Smin + \Delta S_j, \quad (8)$$

where $\Delta S_j = (Smax - Smin) * random_j(0,1)$. $Smin$ and $Smax$ are minimal and maximal possible values of the sensibility. $Smin = Pmin$, $Smax = Pmax$. $Pmin$ and $Pmax$ are minimal and maximal possible values of the pheromone marks. The process continues with selection of a start location for a new exploratory walk. The ability for decision-making based on the

achieved from the exploration (which can be in contradiction with the existing assumptions about the problem during the implementation of the algorithm) supports a good performance across variety of problems, adaptation and self-regulation without additional external adjustments. Selection for a start location x_{0j} for an exploration walk is:

$$x_{0j} = x_k (P_k \geq S_j), \quad (9)$$

where $j = 1, \dots, m$, j is the individuals number; $k = 1, \dots, m$. k is the location number marked with pheromone; x_{0j} is the start location selected from animal number j .

After the exploration follows termination. Acceptable criteria for termination are:

- reaching the optimisation criteria: $f_{max} \geq f_{opt}$,

where f_{max} is the maximal achieved solution, f_{opt} is an acceptable value of the objective function;

- expiration of the generation limit: $g \geq G$,

where G is the limit and g - current value;

- complex criterion: $((f_{max} \geq f_{opt}) \parallel (g \geq G))$.

A specific original peculiarity of Free Search, which has no analogue in other evolutionary algorithms, is a variable called sense. It can be likened as a quantitative indicator of the sensibility. The algorithm tunes the sensibility during the process of search as function of the explored problem. The same algorithm makes different regulations of the sense during the exploration of different problems. This is considered to be a model of adaptation [12].

The presence of variable sense distinguishes individuals from solutions. The individuals are search agents differentiated from the explored solutions and detached from the problems' search space. A solution in FS is a location from a continuous space marked with pheromone. The individuals explore, select, evaluate and mark these solutions.

An individual in FS can be described by the abstraction – an entity, which can move and can evaluate (against particular criteria) locations from the search space thereby indicating their quality. The indicators can be interpreted as a record of previous activities. The individual can identify the pheromone marks from previous activities and can use them to decide where and how to move. It is assumed that all these characteristics are typical of the manner in which animals behave in nature. Therefore the individuals in Free Search are called animals. The variable sense when considered in conjunction with the pheromone marks can be interpreted as personal knowledge, which the individual uses to decide where to move.

The variable sense plays the role of a tool for regulation of divergence and convergence within the search process and a tool for guiding the exploration [12]. A consideration of three idealised general states of sensibility distribution can clarify its self-regulation. These are – uniform, enhanced and reduced sensibility. The relation between sensibility and pheromone distribution affects the decision-making policy of the whole population.

In case of uniformly distributed sensibility and pheromone (Figure 2), the individuals with low level of sensibility can select for start position any location marked with pheromone. The individuals with high sensibility can select for start position locations marked with high level of pheromone and will ignore locations marked with low level of pheromone.



Figure 1: Uniform sensibility

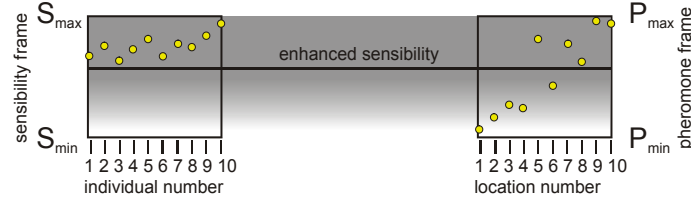


Figure 2: Enhanced sensibility

It is assumed that during a stochastic process within a stochastic environment any deviation could lead to non-uniform changes of the process. The achieved results play a role of deviator. The enhancement of the sensibility urges the individuals to search around the area of the best-found solution from all individuals marked with highest amount of pheromone. This situation appears naturally when the pheromone marks are very different and stochastic generation of the sensibility produces high values. External adding of a constant or a variable to the S_{min} could make an enforced enhancement of the sensibility (Figure 3).

All the individuals with enhanced sensibility will select and can differentiate more precisely locations marked with a higher level of pheromone and will ignore these indicated with lower level of pheromone.

By reducing the sensibility, the individual can be allowed to explore around locations marked with a low level of pheromone. This situation naturally appears when the pheromone marks are very similar and randomly generated sensibility is low. In this case the individuals can select locations marked with low level of pheromone with high probability, which indirectly will decrease the probability for selection of locations marked with high level of pheromone. Subtracting of a constant or a variable from the S_{max} could make an enforced reduction of the sensibility frame (Figure 3).

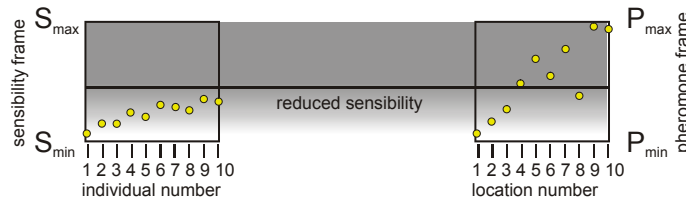


Figure 3: Reduced sensibility

The sensibility across all the individuals varies. Different individuals can have different sensibility: $S_j \neq S_l$ for $j \neq l$, where j and l are numbers of different individuals, $j = 1, \dots, m$, $l = 1, \dots, m$, m is population size. The sensibility varies also during the optimisation process, and one individual can have different sensibility for different explorations.

$S_{jg} \neq S_{jq}$ for $g \neq q$, where j is a current number of an individual, $j = 1, \dots, m$, m is population size, g and q are numbers of different explorations, $g = 1, \dots, G$, $q = 1, \dots, G$, G is the limit of exploration.

The exploration walks begin with selection of start positions. Any location marked with pheromone, which suits the sense of an individual can be selected. The decision relates the sense and the action. This relation could be considered analogous to thought processes. It allows the individual to explore any area of the search space starting from any of the marked locations – the best, the worst or an average.

Free Search performs an adaptive self-regulation of sense, action, and pheromone marks. This adaptive self-regulation is organised as follows. An achievement of better solutions increases the maximal value of the pheromone P_{max} . An increase of the P_{max} increases the maximal allowed sensibility of the individuals S_{max} . This is an adaptive regulation between pheromone and sensibility. In fact it is an abstract approach for learning. The sensibility can be considered as high-level abstract cognition about the explored space based on the achieved and assessed result. The individuals do not memorise any data or low-level information, which consume computational resources. By using sense they build cognition about the quality of the search space and in the same time create skills how to recognise further, higher or lower quality locations. Cognition and skills are abstracted from the achieved results. From a philosophical point of view, “abstraction is a form of cognition based on separation in thought of essential for particular purpose entities, characteristics and relationships” [12]. The abstracted cognition influences thinking. The thinking defines behaviour and action. In computer modelling, abstraction influences operation and self-organisation of algorithms. The abstracted cognition defines behaviour of computational process and it's functioning. The computational process defines action of the computer system and achieved results [12].

Based on relationship between sense and action Free Search implements a computational model of abstraction, cognition, decision-making and action analogous to the processes of perception, learning and thinking in biological systems. This is implemented in the following manner. The better achievements and the higher level of distributed pheromone support enhancement of the sensibility. A higher sensibility does not restrict or does not limit the abilities for movement. It implicitly regulates the individuals' action in terms of selection of a start location for exploration.

During the exploration walk they continue to do small or large steps according to the modification strategy, without restrictions such as convergence rules. However, enhanced sensibility changes their behaviour. They give less attention to steps or locations, which brings low quality results. They can be attracted with high probability from locations with better quality. If small steps achieve better locations the individuals explore these near locations with higher probability. If large steps achieve better results the individuals explore remote locations with higher probability. In this way sensibility adaptively regulates the action. These regulations can be classified as stochastic and probabilistic. Explicit restrictive rules are not applied. The individuals are allowed to explore any location of the search space and enhanced or reduced sensibility increases or decreases the probability for action. The optimisation process keeps the chances of the algorithm to reach the desired solution anywhere in the space. The experience and knowledge can regulate the probability for particular action less than one and greater than zero but they do not determinate such action with a probability of one (100%) or zero (0%).

2. Visual Tool

Visualization tool for advanced search algorithms (VOTASA) is used for evaluation of selected search algorithms.

The tool generates 3-dimensional visual landscape of selected test functions. Optimization process is animated. Sequentially generated solutions model individuals' “movement” on the function's landscape. A three dimensional Cartesian system is displayed around the function, so the user can have a clear view over dimensions and scale.

3. Visual Analysis

This section illustrates various aspects of visual analysis using screenshots of the search process generated by visualization tool VOTASA.

3.1. Dependence on Initialisation

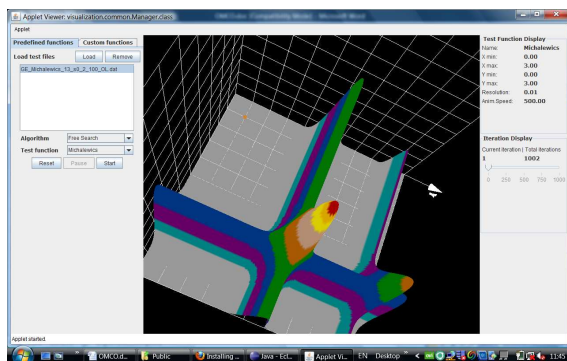


Figure 4: GA start search A from one location

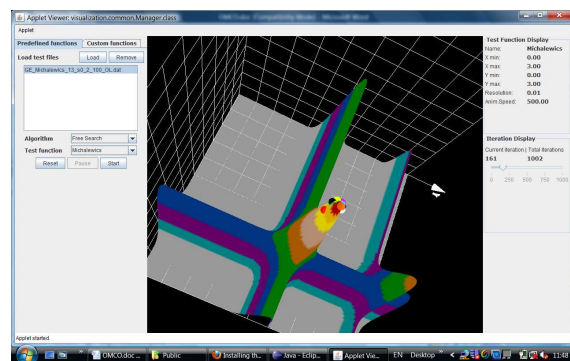


Figure 5: GA end of search A

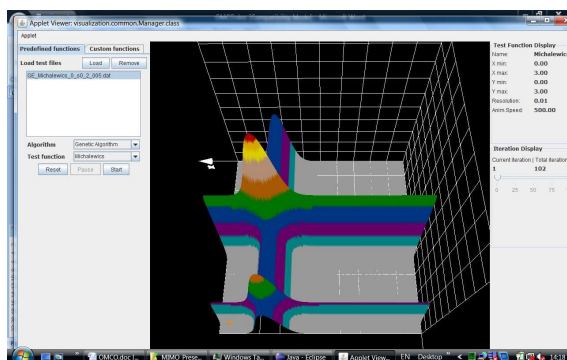


Figure 6: GA start search B from one location

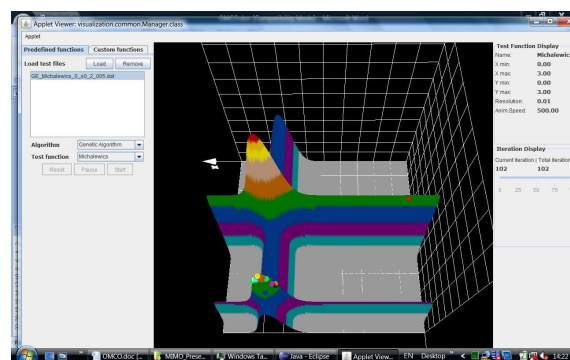


Figure 7: GA end of search B

Figures 4, 5, 6 and 7 illustrate dependence on initialisation. Figure 4 shows search process start form single point appropriately located to the global solution and Figure 5 shows successful end of this process. Figure 6 shows search process start form single point located close to local solution and Figure 7 shows end of this process trapped in local hill.

3.2. Divergence

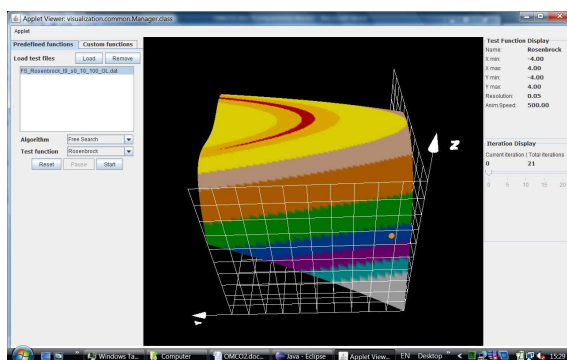


Figure 8: FS start from one location

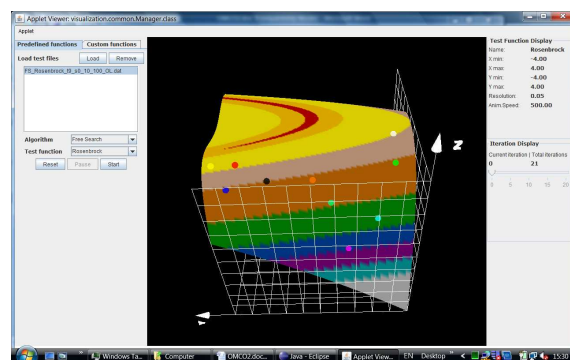


Figure 9: FS divergence within initial walk

Figures 8 and 9 illustrate start form one location and divergence across the search space.

3.3. Convergence

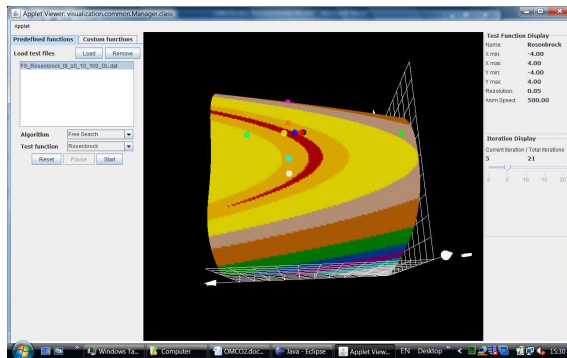


Figure 10: FS approaches maximum

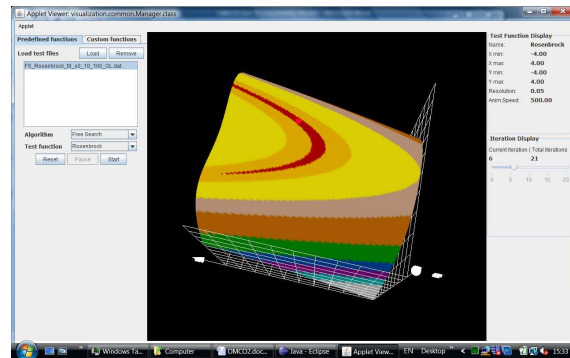


Figure 11: FS reaches maximum

Figures 10 and 11 illustrate how process started on Figure 8 continues with approaching and reaching the maximal solution.

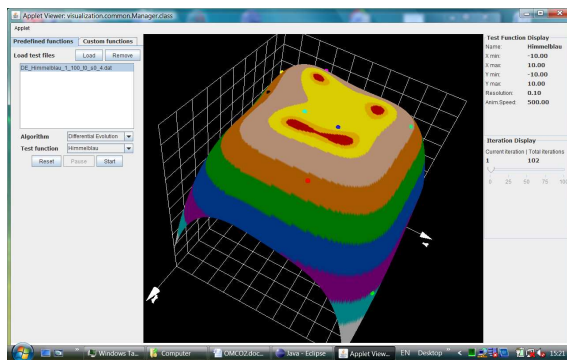


Figure 12: DE start on Himmelblau test

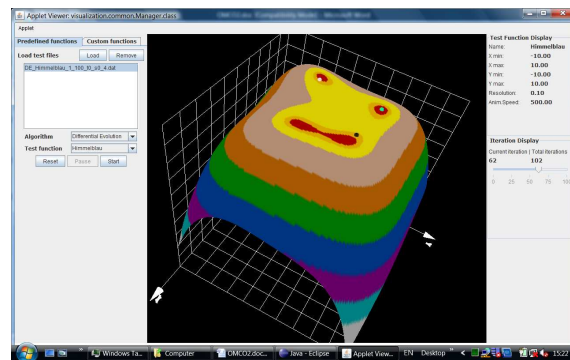


Figure 13: DE converges to optimal solutions

Figure 12 illustrates how Differential Evolution starts from random locations on Himmelblau test, which has four equal value optima.

Figure 13 then shows how Differential Evolution discovers and converges to three of these solutions within 63 iterations.

3.4. Role of Accidental Events

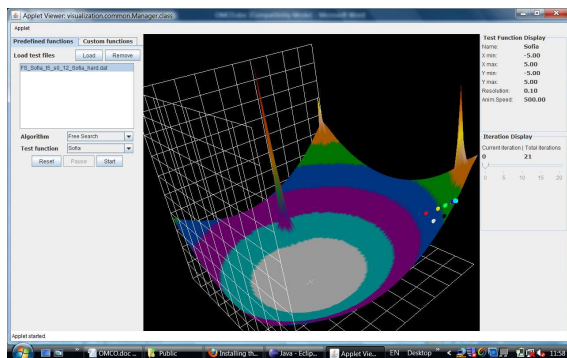


Figure 14: FS start on Sofia test

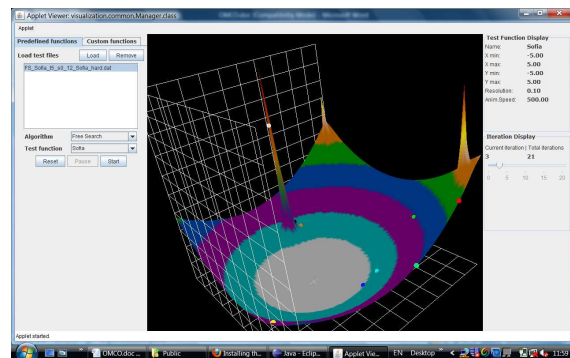


Figure 15: FS accidental event on Sofia test

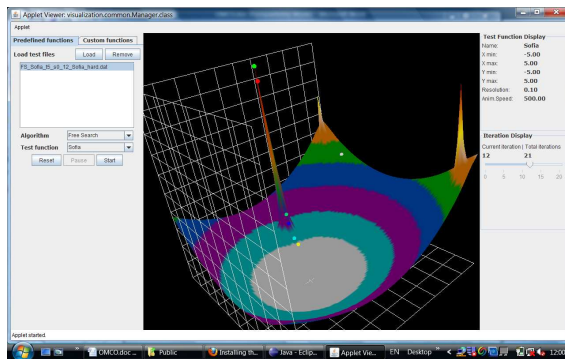


Figure 16: Use of accidental event on Sofia test

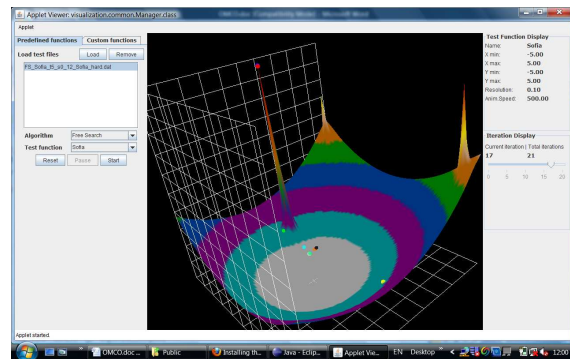


Figure 17: FS achieved maximum on Sofia test

Figure 14 shows Free Search start on Sofia test purposefully selected away from the optimal pack. Gradient on more than 90% of this test is in opposite direction to the maximum. Figure 15 shows generation of accidental even within the area of optimal peak. Figure 16 shows other individuals are attracted within the area, and on Figures 17 is visible achievement of the maximum.

4. Discussion

Visual analysis on dependence on initialisation, convergence, divergence and accidental events role in search process confirms previous studies and suggest new conclusions.

Regarding dependence on initialisation most dependant and sensitive to initial start locations are Differential Evolution and Particle Swarm Optimisation. These methods could converge very fast to the global optimum in at least one initial locations are situated appropriately to the global solution. However if initial locations are away from the area of global solution they face difficulties to identify global area. Once trapped in local hill Particle Swarm Optimisation and Differential Evolution converge quickly to the local solution and have no mechanism to escape. Genetic Algorithm is less dependent on initialisation and could start from single location. For Genetic Algorithm when starts form one location search process starts after first successful mutation. Visual analysis suggests that Genetic Algorithm could escape from trapping if an accidental mutation produce location closes the global solution. However probability for such mutation is very low.

According to the divergence observation suggests that Particle Swarm Optimisation and Differential Evolution could slightly diverge on initial stage of the process. In the middle and in the end of the search process Particle Swarm Optimisation and Differential Evolution shows narrow convergence. Visual analysis confirms that Free Search keeps good divergence abilities during the entire search process.

Regarding convergence visualised search processes clearly confirm that Particle Swarm Optimisation and Differential Evolution have excellent convergence abilities. Genetic Algorithm also demonstrates very good convergence. Free Search has no convergence rule and this is visible. Its ability to discover global solution is based on abstracted knowledge from previous iterations, which reflect on its abilities to avoid trapping and facilitate escaping from trapping.

Observation on how algorithms utilise accidental events confirms that abilities for generation and effective use on accidental events could improve significantly performance. Particle Swarm Optimisation, Genetic Algorithm and Differential Evolution could generate accidentally good solution during initial stage of the search process. Then probability to

generate accidentally remote locations is restricted by their modification strategies to zero. Visualisation of the Free Search process confirms that it is cabala to generate accidentally remote locations during the whole search process. If good location is generates close to the end of the search process precision of the result could be low. However in order to reach better precision, utilising Free Search ability to start from single location, the achieved result could be used for start location for a next run.

Visual analysis helps to identify that Norwegian test [12] has maximum higher than 1.0. For 10 dimensional version of this achieved result is $f_{10} = 1.0000056276962146$ and corresponding variables are presented in Table 1.

Table 1. Norwegian test - variables for 10 dimensional result

x0 = 1.0001125410314771	x1 = 1.0001125413709611	x2 = 1.0001125411045073
x3 = 1.0001125410947156	x4 = 1.0001125411117688	x5 = 1.0001125409284484
x6 = 1.0001125411975889	x7 = 1.0001125410082294	x8 = 1.0001125410497957
x9 = 1.0001125411752061		

Visual analysis shows how balance between divergence and convergence helps to resolve successfully global optimisation problems. Uncertainty in Free Search supports the balance between divergence and convergence and in fact excludes typical for majority Evolutionary algorithms dilemma – Exploration versus Exploitation [7] where algorithms are unable to abstract knowledge from current search process or to utilise this knowledge if exists to improve further behaviour.

Good examples for abstraction of knowledge from the search process and utilisation of this knowledge are numerical tests where optimal value is unknown such as Bump test [8]. When optimum is unknown selection of appropriate initial location is difficult or impossible. This highly applies for real world tasks and optimisation problems. So that search algorithms, which are dependent on initialisation heed special positioning of initial population without any guarantee. Initialisation becomes even harder when the objective function variables number is high. In such cesses relation of large population increases period of search and for time consuming objective functions search process becomes infeasible.

In distinction from these methods Free Search does not depend on initialisation. It could start form one location and diverge in few steps across the search space. This is visible form Figures 8, 9, and 10. In contrast to stochastic search for appropriate initial position Free Search abstracts knowledge from explored accidental (stochastically) locations, then learns this knowledge and use it to improve its further behaviour [10]. These abilities are best visible on Figures 14, 15, 16 and 17.

As additional illustration on multidimensional search space, which cannot be visualised Free Search is tested on 200 dimensional version of Bump test [8]. Achieved result in June 2012 before OMCO NET - 2012 conference is: $f_{200} = 0.85066363874546513$. Constraint for this value is: $p_x = 0.75000000001700473$.

Corresponding to this result variables are presented in Table 2.

This result could be used for comparative assessment of other methods. It will be a challenge to see variables, which produce better solution. Methods which depend on initialisation and relay on initial knowledge could be used for such initial knowledge presented in Table 1 variables.

Table 2. Bump test variables for 200 dimensional results

x0=9.4317265092875271	x1=9.4210880763449794	x2=6.2882470551006726	x3=6.2812044305775245
x4=6.2741502992387712	x5=6.2671333083052971	x6=3.1689464500548583	x7=3.1654470461737678
x8=3.1619395428821764	x9=3.1584654487855133	x10=3.1549887585381216	x11=3.1515052609245462
x12=3.1480514316348942	x13=3.144592385248107	x14=3.1411665208632975	x15=3.1377256453531199
x16=3.1342874197299722	x17=3.1308682053155081	x18=3.1274554202686256	x19=3.1240551403195638
x20=3.1206397745048395	x21=3.1172452805912063	x22=3.1138450849729966	x23=3.1104547305099155
x24=3.1070832311300736	x25=3.1036959150231214	x26=3.1003096268012658	x27=3.0969346627763605
x28=3.0935597373930599	x29=3.0901843874090842	x30=3.0868198721408615	x31=3.0834499576245431
x32=3.0800848219433044	x33=3.0767286653290187	x34=3.0733526058628713	x35=3.0699987823771764
x36=3.0666258516333507	x37=3.0632720102951976	x38=3.0598966938543795	x39=3.056530747298678
x40=3.0531402834188994	x41=3.0497919194950507	x42=3.0464144290947432	x43=3.0430407269387443
x44=3.0396564701050011	x45=3.0362606850701988	x46=3.0328889354156199	x47=3.0294878068600708
x48=3.0260868470445264	x49=3.0226970898753449	x50=3.0192938974016523	x51=3.0158656704071416
x52=3.0124692389801191	x53=3.0090220962794261	x54=3.0055937860424233	x55=3.0021473589341481
x56=2.9987124259525304	x57=2.9952468184703789	x58=2.9917779650003822	x59=2.9883074400256104
x60=2.9848201306512165	x61=2.9813333782210765	x62=2.9778122424860944	x63=2.9743090676481345
x64=2.9707730736866949	x65=2.9672348176413172	x66=2.9636847955273784	x67=2.9601062040242381
x68=2.9565227981730566	x69=2.9529188126611881	x70=2.949321676614483	x71=2.9456856440913848
x72=2.9420258470166796	x73=2.9383818720050225	x74=2.9347045145772901	x75=2.9309963260492711
x76=2.9272779930565633	x77=2.9235334986374415	x78=2.9197768976239793	x79=2.9159920376725164
x80=2.9122020062441676	x81=2.9083602523092282	x82=2.9045094322369045	x83=0.45823984817375735
x84=0.45774089710037869	x85=0.45727114217197029	x86=0.4567745894309917	x87=0.4562951728187824
x88=0.45580577745518042	x89=0.45534177489250671	x90=0.45486045054327734	x91=0.45439602845627802
x92=0.45394310840640006	x93=0.4534702935383077	x94=0.4529999847765222	x95=0.45253467706979733
x96=0.45205813151398533	x97=0.45163461981247027	x98=0.45116993319532678	x99=0.4507212587173699
x100=0.45025372360371446	x101=0.44980480403796747	x102=0.44935699125401113	x103=0.44892267440206884
x104=0.44846630677602395	x105=0.44803663148901179	x106=0.44760106190153104	x107=0.44716738095279651
x108=0.44669963065345109	x109=0.446274123158933	x110=0.44585074831078803	x111=0.44539938407422552
x112=0.44498965251514611	x113=0.44458571277670311	x114=0.44413350683067765	x115=0.44371510577426843
x116=0.44329443469773994	x117=0.44286821129784115	x118=0.44245267787636511	x119=0.44205484519500648
x120=0.44163007466742993	x121=0.44120855371288265	x122=0.44081135789805459	x123=0.44038286798602383
x124=0.43998590163931056	x125=0.43958017705749719	x126=0.43916474548458379	x127=0.43876240741271488
x128=0.43837501452723415	x129=0.43796220309859185	x130=0.43757082525431978	x131=0.43718316066159979
x132=0.43678232263091576	x133=0.43638572567134154	x134=0.43601937796054058	x135=0.43560927575532765
x136=0.43521481062800438	x137=0.43484601432010311	x138=0.43444614837081474	x139=0.43405866650450076
x140=0.43369265901309439	x141=0.43330904830153699	x142=0.4329369251673581	x143=0.43254129852963724
x144=0.43215764564206566	x145=0.43179115637484472	x146=0.43140938352389718	x147=0.43103236473890894
x148=0.43067293713980515	x149=0.43031035434332254	x150=0.42992659626446544	x151=0.42956155047094335
x152=0.42918633562991876	x153=0.42881073599387187	x154=0.42846587868984493	x155=0.42809489894581787
x156=0.42772868753273241	x157=0.42738075957025207	x158=0.42704045925640788	x159=0.426657402002658
x160=0.42631824711613514	x161=0.42594983416198534	x162=0.42561228588155675	x163=0.42524880923746006
x164=0.42489999598120565	x165=0.42454938919611024	x166=0.42420793755776337	x167=0.42385506320185307
x168=0.42352014710422675	x169=0.42316372587208445	x170=0.42281936413544152	x171=0.42247821548464565
x172=0.42214409184863272	x173=0.42180508824841179	x174=0.42145599428238661	x175=0.42110540808126568
x176=0.42076842458400371	x177=0.42043315088331829	x178=0.42010917271001941	x179=0.41979625020274952
x180=0.41946151902511214	x181=0.41911049740812506	x182=0.41878042265486326	x183=0.41845325638208147
x184=0.41813241449989075	x185=0.41781316082099201	x186=0.41746717297257324	x187=0.41716175933695726
x188=0.41681473517720607	x189=0.41651618484573272	x190=0.41617618429311293	x191=0.4158466491643748
x192=0.41553632974871946	x193=0.41520773200359601	x194=0.41489921822492909	x195=0.41456812526745412
x196=0.41424967110372368	x197=0.41393940653707634	x198=0.41362772347405607	x199=0.41332913145585537

5. Conclusion

In summary this article points out potential benefits of visual analysis of Real-Coded Genetic Algorithm, Particle Swarm Optimisation, Differential Evolution and Free Search applied to global optimisation numerical test. Used Visualization tool for advanced search algorithms (VOTASA) shows numerical test as 3D graphics landscape and animates entire search process. This facilitates study and understanding of essential issues such as dependence on initialisation, divergence across the whole search space, convergence to optimal solution, use of accidental events and abilities to abstract knowledge appropriate for performance improvement. Figures in 3D graphics illustrate in certain extent usability of the tool and potential benefit for global optimisation tasks where stagnation in suboptimal solutions is common problem for many methods.

Further research could focus on visual evaluation of other methods and integration with computer aided systems which rely on optimisation.

Acknowledgement

Preparation of this paper is supported by Southampton Solent University, Research and Enterprise Fund, Grant 516/17062011.

References

1. Angeline P., 1998, Evolutionary Optimisation versus Particle Swarm Optimisation: Philosophy and Performance Difference, The 7-th Annual Conference EP, San Diego, USA.
2. Eberhart R. and Kennedy J., 1995, Particle Swarm Optimisation, Proceedings of the 1995 IEEE International Conference on Neural Networks, Vol. 4, 1942-1948.
3. Eberhart R., and Shi Y., 1998, Comparison between Genetic Algorithms and Particle Swarm Optimisation, The 7-th Annual Conference on Evolutionary Programming, San Diego, USA.
4. Eshelman L. J., and Schaffer J. D., 1993, Real-coded genetic algorithms and interval-schemata, Foundations of GA 2, Morgan Kaufman Publishers, San Mateo, pages (187-202).
5. Goldberg D.E., 1989, Genetic Algorithms in Search, Optimisation, and Machine Learning, Addison Wesley Longman Inc. ISBN 0-201-15767-5.
6. Hedar A.R., 2010, Global Optimisation, Kyoto University, http://www-optima.amp.i.kyoto-u.ac.jp/member/student/hedar/Hedar_files/TestGO_files/Page2376.htm, last visited 02.06.10
7. Holland J., 1975, Adaptation in Natural and Artificial Systems, Uni. of Michigan Press.
8. Keane A. J., 1996, A Brief Comparison of Some Evolutionary Optimization Methods, In V.Rayward-Smith, I. Osman, C. Reeves and G.D. Smith, J. Wiley (Editors), Modern Heuristic Search Methods, ISBN 0471962805, pp. 255-272.
9. Penev K., and Littlefair G., 2005, Free Search – A Comparative Analysis, Information Sciences Journal, Elsevier, Vol. 172, Issues 1-2, pp. 173-193.
10. Penev K., 2008, Free Search of Real Value or How to Make Computers Think, Alexander Gegov (Editor), St. Qu publisher, April 2008, ISBN 978-0955894800, UK.
11. Price K., and R. Storn, 1997, "Differential Evolution", Dr. Dobb's Journal, Vol. 22(4), pp. 18-24.
12. Shi Y. and Eberhart R. C., 1998, Parameter Selection in Particle Swarm Optimisation, Evolutionary Programming VII (1998), LNCS, Vol. 1447, pp. 591-600, Springer.
13. Storn R. and Price K., 1995, Differential Evolution – A simple and efficient adaptive scheme for global optimisation over continuous spaces, TR-95-012, International Computer Science Institute, 1947 Center Street, Berkeley, CA 94704-1198, Suite 600.
14. Weisstein E., Global Optimization, WolframMathWorld, <http://mathworld.wolfram.com/GlobalOptimization.html> last visit 02.06.12
15. Wolpert D.H., and W.G. Macready, 1997, No Free Lunch Theorems for Optimisation, IEEE Trans. Evolutionary Computation, Vol. 1(1), pp. 67-82.

Free Search and Particle Swarm Optimisation applied to Non-constrained Test

Vesela Vasileva¹, Kalin Penev²

Technology School, Southampton Solent University, UK

¹Vesela.Vasileva@solent.ac.uk, ²Kalin.Penev@solent.ac.uk

Abstract: This article presents an evaluation of Particle Swarm Optimisation (PSO) with variable inertia weight and Free Search (FS) with variable neighbour space applied to non-constrained numerical test. The objectives are to assess how high convergence speed reflects on adaptation to various test problems and to identify possible balance between convergence speed and adaptation, which allows the algorithms to complete successfully the process of search on heterogeneous tasks with limited computational resources within a reasonable finite time and with acceptable for engineering purposes precision. Modification strategies of both algorithms are compared in terms of their ability for search space exploration. Five numerical tests are explored. Achieved experimental results are presented and analysed.

Keywords: Optimisation, Free Search, Particle Swarm Optimisation, Convergence Speed, Adaptive Methods.

1. Introduction

Investigation on tuning and improvement of convergence speed on various optimisation and search methods attracts research efforts [1][5][8][15][18]. However according to other publications high convergence speed usually increases probability for trapping in suboptimal solutions [7][13][17]. In order to identify a balance between high convergence speed and low probability for trapping this study focuses on Particle Swarm Optimization (PSO) and Free Search (FS). Low probability for trapping usually refers to the algorithms abilities for adaptation. When an algorithm can adapt to various task without changes of its search parameters is could be determined as adaptive algorithm [13]. Recent work investigates how different algorithms operate over different test problems in a limited time and also represents the importance of comparison between different methods' performance. Their execution may vary greatly whether they are applied to hard test problems [10]. For reducing the risk of uncertainties precise tests and analyses are made in terms of their ability for search space exploration [10].

Critical element for balanced search and good capabilities for any adaptation is modification strategy. For some algorithms modification strategy implicitly determines restrictions, which lead to low abilities for adaptation. Such algorithms require tuning of optimisation parameters for each particular task and if the method is applied to other task with the same parameters settings it cannot achieve acceptable results.

This study analyses modification strategy of PSO and FS and points how these strategies could be used in support of balanced set of search parameters. Then PSO and FS are applied to five well-known from the literature numerical optimisation problems generalised for multidimensional optimisation [6][13][18]. For fear assessment of the results each test is modified to ten dimensional variant.

1.1. Particle Swarm Optimization (PSO)

PSO is a classical algorithm used for search and optimisation [10]. Various modifications are published [9]. According to some publications PSO intends to model a social behaviour of a group of individuals whether it searches gradually for the optimum changing the values of the set of solutions [13]. However observation of process of search generated by PSO [18] suggests that its behaviour could be like self-organised particles in cloud systems.

Each particle (individual) shows a single intersection of all search dimensions and is defined as a potential solution to a test problem in multi-dimensional space. The particles appraise their position relative to an objective function (fitness) at every iteration whether particles in a local neighbourhood allocate memories of their best positions then use those memories to accommodate their own velocities, and thus positions [13]. Original concept is modified by adding inertia factor for velocities tuning [3][4]. This study uses modified PSO with variable inertia factor proposed earlier [13].

The velocity v is used to compute a new position for the particle as shown below:

$$x'_{id} = x_{id} + v_{id} \quad (1)$$

where x'_{id} is new position of particle i for dimension d , x_{id} is its current position and v_{id} is its velocity. The velocity vector v'_{id} for each particle is calculated using the best particles' achievement g_d , best for all population achievement P_{id} and inertia factor w according to the equation below:

$$v'_{id} = w * v_{id} + n_1 * \text{random}(0,1) * (P_{id} - x_{id}) + n_2 * \text{random}(0,1) * (g_d - x_{id}) \quad (2)$$

Whether the constants n_1 (individual learning factor) and n_2 (social learning factor) are usually set with the equal values in terms of giving each component equal weight as the cognitive and social learning rate.

Both velocity component and inertia factor support adaptation to the explored test problem. PSO could be adjusted easily as it contains a few parameters only.

1.2. Free Search

Free Search could be best described as adaptive heuristic. This section refines the description published earlier [11], and aims to illustrate the manner in which a computational program can model processes that could be considered similar to thinking and reasoning. FS generates a new solution as deviation of a current one:

$$x = x_0 + \Delta x, \quad (3)$$

where x is a new solution, x_0 is a current solution and Δx is modification strategy. Other interpretation of Δx is that this is simply individuals' step. Individuals in FS explore the search space walking step by step. x , x_0 and Δx are vectors of real numbers. The modification strategy used in the algorithm is calculated according to the equation below:

$$\Delta x_{tji} = R_{ji} * (X_{\max i} - X_{\min i}) * \text{random}_{tji}(0,1), \quad (4)$$

where i indicates dimension; $i = 1, \dots, n$ for a multi-dimensional step; n is dimensions number; t is the current step $t = 1, \dots, T$. T is the step limit per walk; R_{ji} indicates the size of the idealised frame of the neighbourhood space for individual j within the dimension i . $\text{random}_{tji}(0,1)$ generates random values between 0 and 1. Δx_{tji} indicates the actual size of the step for step t of individual j within dimension i .

During the exploration an individual with a neighbourhood space, which exceeds search space boundaries, can perform global exploration whereas another individual with small neighbourhood space can make precise steps around one location.

The modification strategy is independent from the current or the best achievements and this is fundamental difference from PSO. The exploration performs heuristic trials based on stochastic divergence from one location. The concrete value of the neighbourhood space for a particular exploration defines the extent of uncertainty of the chosen individual. The exploration walk is followed by an individual assessment of the explored locations. The best location is marked with pheromone. The pheromone indicates the quality of the locations

and may be considered as a result or cognition from previous activities. The assessment, during the exploration, is defined as follows:

$$f_{ij} = f(x_{ij}), f_j = \max(f_{ij}), \quad (5)$$

where f_{ij} is the value of the objective function achieved from animal j for step t . f_j is the quality of the location marked with pheromone from an individual after one exploration. The pheromone generation is generalised for the whole population:

$$P_j = f_j / \max(f_j), \quad (6)$$

where $\max(f_j)$ is the best achieved value from the population for the exploration.

This is a normalisation of the explored problem to an idealised qualitative (or perhaps cognitive) space, in which the algorithm operates. This idealised space uses for a model an idealised space of notions in thought of biological systems, in which they generate decisions. The normalisation of any particular search space to one idealised space supports adaptation and successful performance across variety of problems without additional external adjustments. The sensibility generation is:

$$S_j = S_{min} + \Delta S_j, \quad (7)$$

where

$$\Delta S_j = (S_{max} - S_{min}) * \text{random}(0,1) \quad (8)$$

S_{min} and S_{max} are minimal and maximal possible values of the sensibility.

$S_{min} = P_{min}$, $S_{max} = P_{max}$.

P_{min} and P_{max} are minimal and maximal possible values of the pheromone marks. The process continues with selection of a start location for a new exploratory walk. The ability for decision-making based on the achieved from the exploration (which can be in contradiction with the existing assumptions about the problem during the implementation of the algorithm) supports a good performance across variety of problems, adaptation and self-regulation without additional external adjustments. Selection for a start location x_{0j} for an exploration walk is:

$$x_{0j} = x_k (P_k \geq S_j), \quad (9)$$

where $j = 1, \dots, m$, j is the number of the individuals; $k = 1, \dots, m$, k is the number of the location marked with pheromone; x_{0j} is the start location selected from animal number j . After the exploration follows termination.

A specific original peculiarity of Free Search, which has no analogue in other evolutionary algorithms, is a variable called sense. It can be likened as a quantitative indicator of sensibility. The algorithm tunes the sensibility during the process of search as function of the explored problem. The same algorithm makes different regulations of the sense during the exploration of different problems. This is considered to be a model of adaptation [13]. The variable sense distinguishes the individuals from the solutions. The individuals are search agents differentiated from the explored solutions and detached from the problems' search space. A solution in FS is a location from a continuous space marked with pheromone. The individuals explore, select, evaluate and mark these solutions.

An individual in FS can be described by the abstraction – an entity, which can move and can evaluate (against particular criteria) locations from the search space thereby indicating their quality. The indicators can be interpreted as a record of previous activities. The individual can identify the pheromone marks from previous activities and can use them to decide where and how to move. It is assumed that all these characteristics are typical of the manner in which animals behave in nature. Therefore the individuals in FS are called animals. The

variable sense when considered in conjunction with the pheromone marks can be interpreted as personal knowledge, which the individual uses to decide where to move. The variable sense plays the role of a tool for regulation of divergence and convergence within the search process and a tool for guiding the exploration [11].

2. Modification Strategies Comparison

Comparing modification strategies of PSO and FS it can be identified that inertia weight in PSO plays the same role as neighbour space in FS. Both of them are absolute values, which reflect on convergence speed, abilities for adaptation. Essential difference is that in PSO the current best achievements for the population and for the particles explicitly determine generation of new velocity and then the new individuals. The locations, within the area of the possible velocity values around an individual, have non-zero probability for access. However, for the rest of the search space the probability for access is zero. Increasing inertia factor increases the area with nonzero probability for access. This decreases probability for trapping however this critically decreases convergence speed. Decreasing inertia factor directly decreases the area with nonzero probability for access. This increases the convergence speed to the located peak. However, if this is a local peak PSO has no mechanism to escape. In FS mechanism for trapping avoidance is based on individuals' sense.

3. Numerical Tests

3.1. Griewank Test Function

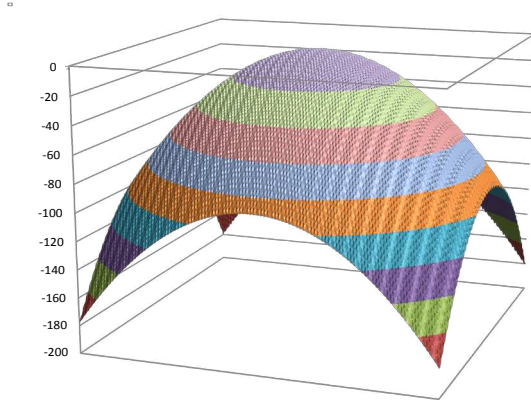


Figure 1: Griewank test

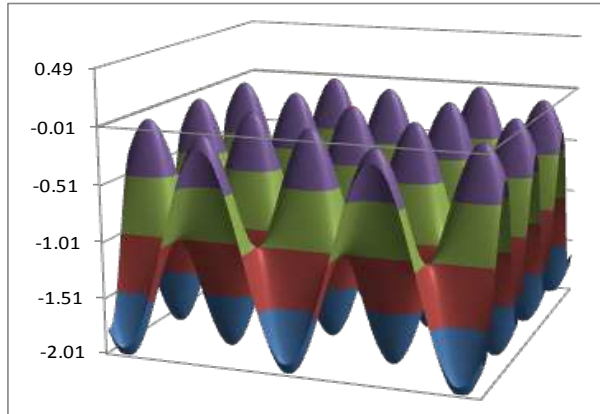


Figure 2: Griewank test – top of the hill

$$f(x_i) = 1 - \frac{1}{4000} \sum_{i=1}^n x_i^2 \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) \quad (10)$$

where $x_i \in [-600.0, 600.0]$. Global solution is $f(0.0, 0.0) = 0.0$.

3.2. Norwegian Test Function

$$f(x_i) = \prod_{i=1}^n \left(\cos(\pi x_i^3) \left(\frac{99 + x_i}{100} \right) \right) \quad (11)$$

where $x_i \in [-1.1, 1.1]$. Global solution is $f(1.00011, 1.00011) = 1.00000113$. All local peaks are below 0.99.

3.3. Rastrigin Test Function

$$f(x) = nA - \sum_{i=1}^n (x_i^2 - A \cos(2\pi x_i)) \quad (12)$$

where $x_i \in [-5.12, 5.12]$. Global solution is $f(0.0, 0.0) = 0.0$.

□

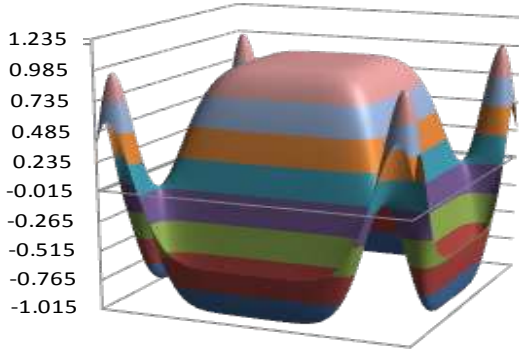


Figure 3: Norwegian test

□

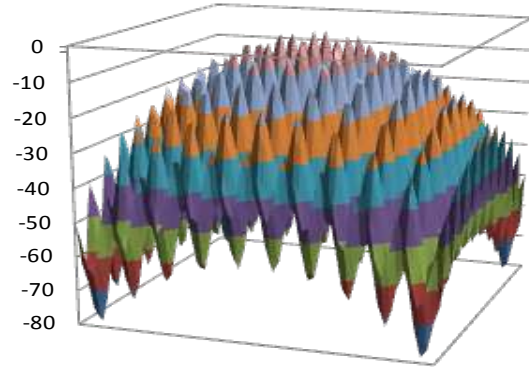


Figure 4: Rastrigin test

3.4. Rosenbrock Test Function

$$f(x_i) = \sum_{i=1}^{n-1} [100 * (x_{i+1} - x_i^2)^2 + (x_i - 1)^2] \quad (13)$$

where $x_i \in [-500, 500]$. Global solution is $f(1.0, 1.0) = 0.0$. The function is shown on Figure 5.

3.5. Sphere Test Function

$$f(x_i) = -\sum_{i=1}^n x_i^2 \quad (14)$$

where $x_i \in [-512, 512]$. Global solution is $f(0.0, 0.0) = 0.0$.

□

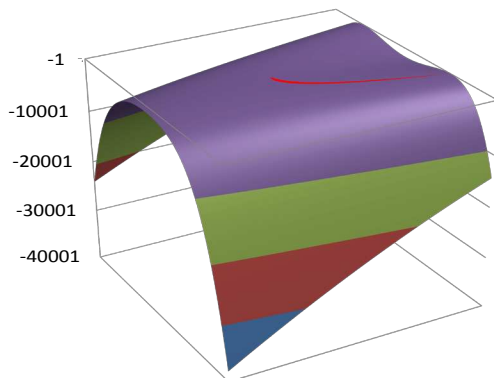


Figure 5: Rosenbrock test

□

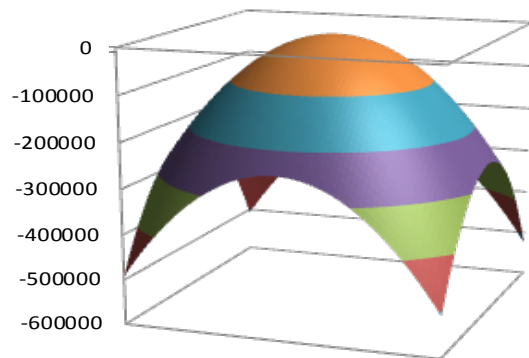


Figure 6: Sphere numerical test

4. Experimental Results

All tests are evaluated for 10 dimensions. For all experiments for both FS and PSO population is 10 individuals. PSO experiments are limited to 2000 and 20000 iterations per each test. Total number of test function evaluations for 10 individuals are accordingly: $2000 \times 10 = 20000$ and $20000 \times 10 = 200000$.

FS experiments are limited to 400 and 4000 explorations with 5 steps per exploration. Total number of test function evaluations for 10 individuals are: $400 \times 5 \times 10 = 20000$ and $4000 \times 5 \times 10 = 200000$.

For all test PSO is applied with the variable inertia factor which enhances to some extent its ability for adaptation. Individual learning factor is 2.0. Group learning factor is 2.0. Inertia weight varies within the interval 0.1 - 1.0 with step 0.1. Initialisation is stochastic:

$$x_{0ji} = Xmin_i + (Xmax_i - Xmin_i) * random_{ji}(0, 1), \quad (15)$$

where $Xmax_i - Xmin_i$ are search space boundaries; j is individual, i is dimension.

For all test FS is used with its standard set of parameters – population size – 10; steps per exploration 5. In order to provide equal conditions for testing with PSO neighbour space varies within the interval 0.1 - 1.0 with step 0.1. Sensibility randomly varies within the interval 0.99999 - 1.0.

To reduce probability for dependence on initialisation, 32 experiments with different initialisation per each inertia value are completed for both FS and PSO. This corresponds to 320 experiments per test per method in total. FS is evaluated additionally to the same number of experiments but with start for all individuals from a single location purposefully selected away from the global optimum. The single location for all tests is defined as:

$$x_{0ji} = Xmin_i + 0.1 * (Xmax_i - Xmin_i), \quad (16)$$

where $Xmax_i - Xmin_i$ are search space boundaries; j is individual; i is dimension.

Achieved best values from 320 experiments are presented in the tables below.

Table 1: PSO start from random locations		
Function	Iterations	
	2000	20000
Rastrigin	-4.9748	-2.98488
Griewank	-0.009857	-0.007396
Rosenbrock	-0.00115796	-0.0023069
Norwegian	0.950406	0.951045
Sphere	-2.9557E-078	0.000000

Table 2: FS start from random locations		
Function	Iterations	
	2000	20000
Rastrigin	-0.000129	-1.4571E-06
Griewank	-0.002767	-1.3650E-06
Rosenbrock	-8.4371E-05	-1.5643E-05
Norwegian	0.966949	0.99998
Sphere	-0.004359	-0.000109

Table 3: FS start from single location		
Function	Iterations	
	2000	20000
Rastrigin	-0.000281	-1.0618E-06
Griewank	-0.000699	-3.08E-06
Rosenbrock	-2.3184E-05	-2.0315E-06
Norwegian	0.980593	0.999974
Sphere	-0.010623	-4.6E-05

PSO demonstrates unapproachable convergence on 10 dimensional Sphere test. It reaches the optimum for less than 2000 iterations. On Griewank, Rosenbrock tests PSO shows dependence on initialisation and accidental events. If initialisation is appropriate it achieves optimal values for ten dimensional variant of these tests within 2000 iterations with precision below 0.01%. Although on Rosenbrock tests from 320 experiments limited to 20000 iterations PSO reaches more results close to the optimum its best result was not better than for 2000 iterations. Identification of reasons for this requires further investigation.

PSO cannot escape from local trap of Norwegian test within 20000 iterations and has difficulties on Rastrigin test. In order to resolve these tests PSO needs retuning of its parameters.

Standard FS configuration has medium convergence speed. iFS [18] is the Free Search modification with highest convergence speed published in the literature. Within 400 explorations (corresponding to 2000 iterations) FS cannot escape from trapping for Norwegian test and cannot reach close to zero value on Rosenbrock test. However for 4000 explorations (corresponding to 20000 iterations) FS resolves all tests for both start from random locations and start from single location. For experiments with start from single location FS confirm its abilities to diverge over the search space and then to identify optimum.

5. Conclusion

Presented study contributes to the knowledge in adaptive computing and heuristic methods in terms of identification of balance between convergence speed and abilities for adaptation for PSO and FS applied to non-constrained global optimisation.

Experimental results confirms: (1) published earlier evaluation that PSO [13] has excellent convergence in local search but is not adaptive enough in global optimization due to constraints produced from the best particles' and all population achievements' P_i and g ; (2) standard FS configuration has balance between convergence (which is medium) and abilities for adaptation and can resolve various problems without retuning of its parameters within reasonable amount of time and with acceptable for engineering purposes precision.

Acknowledgement

Preparation of this paper is supported by Southampton Solent University, Research and Enterprise Fund, Grant 516/17062011.

References

1. P. Angeline, 1998, Evolutionary Optimisation versus Particle Swarm Optimisation: Philosophy and Performance Difference, The 7-th Annual Conference on Evolutionary Programming, San Diego, USA.
2. R. Eberhart and J. Kennedy, "Particle Swarm Optimisation", Proc. Int'l Conf. Neural Networks, Vol. 4, pp. 1942-1948, 1995.

3. R. Eberhart and Y. Shi, "Comparison between Genetic Algorithms and Particle Swarm Optimisation", Proc. Seventh Annual Conf. Evolutionally Programming, San Diego, USA, pp. 611-616, 1998a
4. R. C. Eberhart and Y. Shi, "Parameter Selection in Particle Swarm Optimisation", Evolutionary Programming VII, Lecture Notes in Computer Science, Vol. 1447, pp. 591-600, 1998b.
5. S. Fidanova, Convergence Proof for a Monte Carlo Methods for Combinatorial Optimization Problems, Computer Science-ICCS-2004, Lecture Notes in Computer Sciences No 3039 ISBN: 3-540-22129-8, Springer, Germany, 2004, 523-534.
6. A. O. Griewank, "Generalized Decent for Global Optimization", J. Opt. Th. Appl., Vol. 34, pp. 11-39, 1981.
7. C. Igel, M. Toussaint, 2004, A No-Free-Lunch Theorem for Non-Uniform Distribution of Target Functions, Journal of Mathematical Modelling and Algorithms, Vol. 3, pp 313-322, 2004.© 2004 Kluwer Academic Publishers.
8. M. Jiang, Y. P. Luo, and S. Y. Yang, "Stochastic convergence analysis and parameter selection of the standard particle swarm optimization algorithm", Information Processing Letters, Vol. 102, No. 1, pp. 8-16, 2007.
9. L. Lin, Q. Luo, J.-Y Liu, C. Long, "An Improved Particle Swarm Optimization Algorithm", Int'l Conf. Business Management and Electronic Information (BMEI), pp. 838-842, 2011.
10. F. Peng, K. Tang, G. Chen, X. Yao, "Population- Based Algorithm Portfolio for Numerical Optimization", IEEE TRANS Evolutionary Computation, pp. 782-800, 2010.
11. K. Penev, and G. Littlefair, 2005, Free Search – A Comparative Analysis, Information Sciences Journal, Elsevier, Vol. 172, Issues 1-2, pp. 173-193.
12. K. Penev, "Adaptive Intelligence – Essential Aspects", Proc. Int'l Conf. Automatics and Informatics, John Atanasoff Society of Automatics and Informatics, pp. 97-100, 2009.
13. K. Penev, "Free Search of Real Value or How to Make Computers Think", Alexander Gegov (Editor), St. Qu publisher, ISBN 978-0955894800, UK, 2008.
14. H.H. Rosenbrock, "An automate method for finding the greatest or least value of a function", Comput. J, Vol. 3, pp.175-184, 1960.
15. I. C. Trelea, "The particle swarm optimization algorithm: convergence analysis and parameter selection", Information Processing Letters, Vol. 85, No. 6, pp. 317-325, 2003.
16. E. Veliev, and K. Penev, Visualization of Free Search Process. In: Advanced Topics on Evolutionary Computing. WSEAS Press, Bulgaria, pp. 133-135. ISBN 978 960 6766 58 9, 2008.
17. D.H.Wolpert, and W.G.Macready, 1997, No Free Lunch Theorems for Optimisation, IEEE Trans. Evolutionary Computation, Vol. 1(1), pp. 67-82.
18. G.-Y. Zhu, J.-B. Wang, H. Guo, "Research and Improvement of Free Search Algorithm", Int'l Conf. Artificial Intelligence and Computational Intelligence, pp. 235-239, 2009.

Differential Evolution with Enhanced Abilities for Adaptation Applied to Heterogeneous Numerical Optimization

Mariya Lazarova¹, Kalin Penev²

¹Technical University of Sofia, Bulgaria, ²Southampton Solent University, UK

¹mklazarova@gmail.com, ²kalin.penev@solent.ac.uk

Abstract: This article presents an exploration of Differential Evolution (DE) algorithm with enhanced adaptability. The main purpose of this study is to identify how this search method can cope with changes of the number of variables of a hard design test, unaided. The results clearly show that this method successfully solves the explored functions.

Keywords: Differential Evolution, Numerical Optimization.

1. Introduction

In mathematics and computational science, mathematical optimization refers to the selection of a best element from some set of available alternatives [1]. In the simplest case, an optimization problem consists of maximizing or minimizing a real function by systematically choosing input values from within an allowed set and computing the value of the function. More generally, optimization includes a variety of different types of objective functions and different types of domains.

There are different approaches to solve these problems like algorithms that terminate in a finite number of steps, iterative methods that converge to a solution (on some specified class of problems) or heuristics that may provide approximate solution to some problems. In this thesis we will stop our attention specifically on heuristic methods, in particular one of them - Differential Evolution (DE).

Differential Evolution Algorithm is proposed by Kenneth Price and Rainer Storn in 1995 [9]. This is one heuristic approach for optimizing different continuous space functions. This algorithm is a very simple population-based search method and very powerful at the same time. It belongs to the class of stochastic optimization methods. Differential Evolution is applicable to non-linear and non-differentiable functions that can't be solved to traditional approximation techniques. Furthermore it has been successfully applied in diverse fields such as mechanical engineering [10], [7], communication [8] and pattern recognition [2]. The aim of the exploration is to show how DE works and how it can be applied to different test functions, whatever their complexity.

Schwefel function, Schwefel Problem 1.2, Schwefel Problem 2.22 and Branin function are considered in this article. These functions are most often tested for minimum value, but in this study will seek their maximum value by DE algorithm and will perform their graphics. The results suggest that this search method can resolve successfully each considered function independently of the variation of the number of variables and constraint conditions.

2. Conceptual Model of the Differential Evolution Algorithm

Differential Evolution starts with a stochastic selection of an initial set of solutions called design vector. The objective function corresponds to each individual of the population and its value is measure of that individual's fitness as an optimum. The initial population of vectors is transformed generation-by-generation and the best individuals from each generation form a solution vector.

The structure of the DE algorithm likes other population-based search methods. It maintains two arrays, which hold two populations of multi-dimensional real-valued vectors [3]. The

primary array holds the current population [3]. The secondary array accumulates vector which are produced for next generation. Then, the algorithm selects the better vectors from existing vectors or trail vectors. These trail vectors are formed by using mutation and recombination of the vectors in the primary array.

The mutation is an operation that makes small random alternations to one or more parameters of an existing population vector [3]. Differential Evolution uses the population itself as the source of appropriately scaled perturbations. Through this approach, those variable having a narrow and well-defined range around the minimum, will have a small variation among the population members and resulting in their mutations being relatively small [3]. This automatic adaptation significantly improves the behavior of the algorithm and making DE one of the more promising new ideas in optimization.

Differential Evolution performs manipulation with target, donor and differential vectors. The minimal number of vectors in one population has to be more than four. For modification strategies, which use four differential vectors the minimal population size is seven. The current target and the corresponding new trail vector (individual) in each generation are the subject of competitions to determine the composition of the next generation [3]. There are a few steps to generate the new trail vector:

- 1) selection of randomly chosen donor vector from the population, which is different from current target vector;
- 2) selection of randomly chosen differential vectors (two or four), which are different from the donor vector and from current target vector;
- 3) calculation of a difference between differential vectors and scaling it by multiplication with a constant called the differential factors (noted as F) [3];
- 4) adding the difference to the donor vector, which produce a new vector [3];
- 5) crossover between the current target vector and the new vector so that the trail vector inherits parameters from both of them [3].

If the trail vector is better than the current target vector, then the trail vector replaces the target vector in the next generation [3]. There are three factors, which amend evolution of DE-the population size, the scaling weight applied to random differential and the constant that mediates the number of parameters in the crossover operation. They describe DE as a heuristic approach for optimizing non-linear and non-differentiable functions within continuous space [5].

The following indications are introduced:

X_k – the target vector;

X_i and X_j – the differential vectors;

F – the differential factor (weight).

Every pair of vectors (X_i, X_j) in the primary array defines a differential vector $X_i - X_j$. When these two vectors are chosen randomly, their weighted difference is used to perturb another vector in the primary array, X'_k [3]:

$$X'_k = X_k + F(X_i - X_j) \quad (1.1)$$

Other effective variation of this calculation is keeping track of the best vector so far noted as X^* . This vector is combined with X_k and then perturbed that we get the following expression:

$$X'_k = X_k + F(X^* - X_k) + F(X_i - X_j) \quad (1.2)$$

Rainer Storn suggests some modification strategies for calculation of new individual as follows:

$$X'_k = X_k + F(X_i - X_j) \quad (2.1)$$

$$X'_k = X^* + F(X_i - X_j) \quad (2.2)$$

$$X'_k = X_k + F(X^* - X_k) + F(X_i - X_j) \quad (2.3)$$

$$X'_k = X^* + F(X_i - X_j + X_n - X_m) \quad (2.4)$$

$$X'_k = X_k + F(X^* - X_k + X_n - X_m) \quad (2.5)$$

where

X_k – is donor vector;

X'_k – is mutated donor;

X^* – is the best vector for current population;

X_i, X_j, X_n, X_m – are differential vectors;

F – is differential factor.

These strategies (2.1 – 2.5) can be applied to all variables, to part of them or to one variable of the donor vector.

The different strategies for mutation, which are presented by Storn and Price, maintain the influence of the successful member of the population to all trail vectors (strategies 2.2 and 2.4) and generate the differential using more vectors (strategies 2.4 and 2.5).

In the next step each primary array vector X_k is combined with X'_k to produce a trail vector X_t . So the trail vector presents a child of two parents.

Once a new trail solution has been generated, selection determines which among them will survive into the next generation. Each child X_t is pitted against its parent X_k in the primary array [3]. Only the best of the two succeeds to advance into the next generation.

The end of the algorithm occurs when generation limit is expired or the objective function is satisfied. These strategies for generation of a new individual in the Differential Evolution are an original advantage because they make one fast and adaptive algorithm, thus they contribute to the improvement of algorithm for global optimization. The biggest disadvantage of this search method is when the solutions of some functions fall in the trap then they cannot escape from it. Independently of this disadvantage, Differential Evolution Algorithm proves that is one good and powerful search method for global optimization of the functions defined in continuous search space.

3. Methodology of the Exploration

In this study Differential Evolution is implemented in an original variant proposed by Storn. As already mentioned there are three factors by which to change the evolution of this search method and they are same in all test programs.

In our case the number of the initial population is equal to 10 and this size can be flexibly changed. The differential multiplication factor F varies from 0.5 to 1.5 with step 0.1. Crossover operation for generation of the trail vector is implemented as probabilistic replacement of the parameters from the selected vector and parameters from the calculated differential vector added to the donor vector [3]. The probability for this operation is set to 0.5

and it can be changed. In programs of the algorithm the following strategy for mutation is used:

$$X'_k = X_k + F(X^* - X_k) + F(X_i - X_j) \quad (3.1)$$

where the mutation probability is equal to 0.3.

All test functions are developed with the same black-box model of algorithm and with the parameters described above. When testing is not necessary to change the algorithm itself or its parameters irrespective of the different search domain of each test function and their complexity. For this reason the Differential Evolution is an algorithm with enhanced abilities for adaptation applied to heterogeneous numerical optimization.

4. Test Problems

In this exploration are considered four test functions defined in continuous search space. The main objective for all experiments is to find the maximum value of these test functions, therefore they are transformed in relevant manner.

4.1. Schwefel Function

The Schwefel function (Figure 1) is given by the following analytical expression:

$$f_1(x) = 418.9829 * D \sum_{i=1}^D x_i \sin\left(\sqrt{|x_i|}\right) \quad (4.1)$$

The search space for this function is restricted to $-500 \leq x_i \leq 500$, $i=1, \dots, D$, D is the number of dimensions. The maximum value of this test problem for two dimensions is $f_{max}(x^*)=0, x^*=(420.96, 420.96)$.

4.2. Schwefel Problem 1.2

This test function (Figure 2) is given as:

$$f_2(x) = \sum_{i=1}^D \left(\sum_{j=1}^i x_j \right)^2 \quad (4.2)$$

The test problem has the following search range: $-100 \leq x_j \leq 100$. In this equation D is the number of dimensions.

This function is explored for two dimensions and its maximum is $f_{max}(x^*)=0$ for $x^*=(0,0)$.

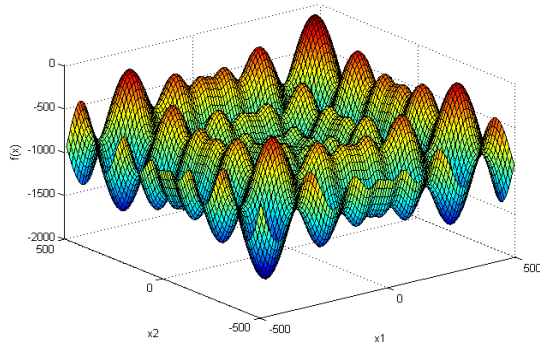


Figure 1: Schwefel function 2D

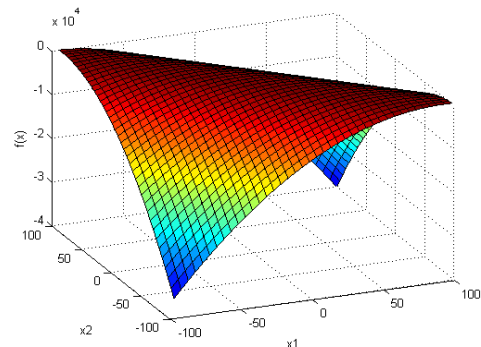


Figure 2: Schwefel Problem 1.2 2D

4.3. Schwefel Problem 2.22

The Schwefel Problem 2.22 (Figure 3) is a global optimization function with the following definition:

$$f_3(x) = \sum_{i=1}^D |x_i| + \prod_{i=1}^D |x_i| \quad (4.3)$$

where search space borders are defined by $-10 \leq x_i \leq 10$, $i = 1, \dots, D$ and D is the number of dimensions. The maximum value of this test problem for two dimensions is $f_{\max}(x^*)=0$ for $x^*=(0,0)$.

4.4. Branin Function

The Branin function (Figure 4.) is a two-dimensional optimization function which is defined as:

$$f_4(x) = \left(x_2 - \frac{5.1}{4\pi^2} x_1^2 + \frac{5}{\pi} x_1 - 6 \right)^2 + 10 \left(1 - \frac{1}{8\pi} \right) \cos x_1 + 10 \quad (4.4)$$

The search domain of this test problem for x_1 is $[-5;10]$ and for x_2 is $[0;15]$. The maximum value is $f_{\max}(x^*)=-0.398$ for $x^*=((-3.142,12.275), (3.142,2.275), (9.425,2.425))$.

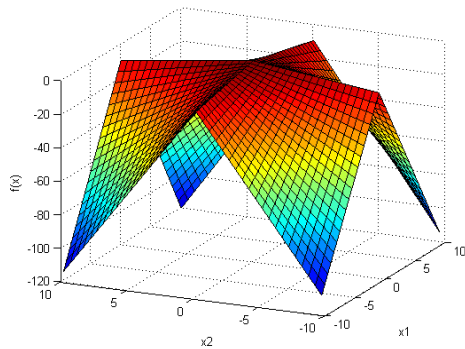


Figure 3: Schwefel Problem 2.22 2D

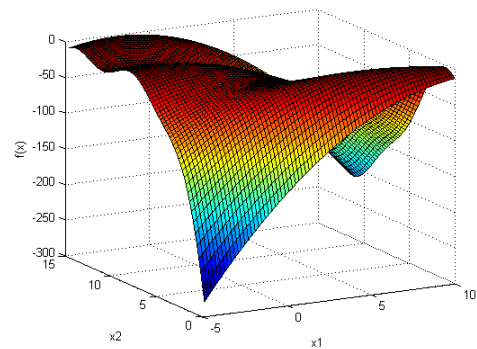


Figure 4: Branin function 2D

5. Experimental Results

Two sets of experiments of 320 evaluations per function are made for 100 and 2000 generations. The results are accepted as successful if: for Schwefel function they are higher than -0.5; for Schwefel Problem 1.2 they are higher than -0.5; for Schwefel Problem 2.22 they are higher than -0.3; for Branin function they are higher than -0.4. All results are showed in the following table:

Table 1: Experimental results				
f	Number of the successful results		Results (%)	
	100G	2000G	100G	2000G
f_1	260	298	81.25%	93.13%
f_2	305	316	95.31%	98.75%
f_3	319	320	99.69%	100%
f_4	309	320	95.56%	100%

6. Conclusion

In this study different test functions are explored, because they are resistant to traditional approximation techniques.

For this purpose whole test problems are developed with the same black-box model of the algorithm. When using this model to solve the function is not necessary to reset the algorithm itself and is not necessary to change the parameters of this algorithm.

Two different series were made for each test function for 100 generations and for 2000 generations. The experimental results show that the average success of the algorithm for 100 generations is 92.95% and for 2000 generations is 97.97%.

Schwefel function, Schwefel Problem 1.2 and Schwefel Problem 2.22 which are tested for two dimensions in this exploration, can be tested for more dimensions. Furthermore other functions can be tested with DE algorithm.

References

1. Dantzig G. B., "The Nature of Mathematical Programming, "*Mathematical Programming Glossary*", INFORMS Computing Society.
2. J. Ilonen, J.-K. Kamarainen, and J. Lampinen, "Differential evolution training algorithm for feed-forward neural networks", *Neural Process. Lett.*, Vol. 7, No. 1, pp. 93-105, 2003.
3. Penev K., 2005. Adaptive search heuristics applied to numerical optimisation, pp.47-51, PhD thesis, SSU.
4. Penev, K., 2006. Free Search and Differential Evolution Towards Dimensions Number Change. In: *Intelligent Engineering Systems Through Artificial Neural Networks*. ASME, New York, USA, pp. 37-42. ISBN 0-7918-0256-6
5. Price K., and Storn R., 1997, Differential Evolution – A simple evolution strategy for fast optimisation, *Dr. Dobbs's Journal*, Vol. 22 (4), pp. 18-24.
6. Qin A., Huang V., Suganthan, 2008, Differential Evolution Algorithm With Strategy Adaptation for Global Numerical Optimization
7. Joshi R. and Sanderson A. C., "Minimal representation multisensory fusion using differential evolution", *IEEE Trans. Syst. Man. Cybern. A, Syst. Humans*, Vol. 29, No. 1, pp. 63-76, Jan. 1999.
8. Storn R., "On the usage of differential evolution for function optimization", *Proc. Biennial Conf. North Amer. Fuzzy Inf. Process. Soc.*, Berkeley, CA, 1996, pp. 519-523.
9. Storm R. and Price K., 1995, Differential Evolution – a simple and efficient adaptive scheme for global optimization over continuous space, ICSI Technical Report TR-95-012
10. Rogalsky, Derksen R.W., and Kocabiyik S. "Differential evolution in aerodynamic optimization", *Proc. 46th Annu. Conf. Can. Aeronaut. Space Inst.*, Montreal, QC, Canada, May 1999, pp. 29-36.

Self-Adapting, Multi-Parent Recombination Evolution Strategy Algorithm

Ivaylo Simeonov¹, Kalin Penev²

¹Technical University, Sofia, Bulgaria, ²Southampton Solent University, UK

¹ivailosim@gmail.com, ²kalin.penev@solent.ac.uk

Abstract: Evolution Strategies (ES) are algorithms similar to Genetic Algorithms (GA) which use the basic principles of natural evolution and adaptation as a method to solve optimization problems [1].

This paper focuses on Self-Adapting, Multi-Parent Recombination Evolution Strategy (MPR-ES) method applied to numerical optimization.

Proposed modified Evolution Strategy Algorithm utilises a new approach in generation of offspring from three randomly chosen individuals. It is also equipped with a mutation strategy in order to bring it out of local optima stuck problems and a self-adaptation replacement strategy ensuring that each generation of individuals is at least as good as the previous one, if not better.

Keywords: Evolution Strategy, Optimization, Self-Adaptation, Evolutionary Computation.

1. Evolutionary Computation Essentials

Evolution Strategy crystallizes in result of research efforts on improving optimization techniques [9][10]. The idea behind evolutionary computation and in particular ES is based on recreating the natural evolution processes. In nature, different individuals reproduce and create offspring of new individuals. The new generation has similarities with the parents but by no means is identical to them. Over time, the offspring generations adapt more and more to the surrounding environment so they become better – they evolve [1][4][6][10].

Sometimes unforeseen mutations happen, which in some cases prove to be useful to the individuals and lead again to evolution.

Also in nature even individuals with less opportunity for reproduction (weaker) get the chance to create offspring. This is not a prerequisite that their descendants will also be weak – in some cases this leads to more advanced successors, therefore weaker individuals should not be ignored in the process of reproduction.

The objective of presented in this paper investigation is to evaluate MPR-ES algorithm on maximization of well-known from the literature global, test functions (also called objective or fitness functions) somehow resistant to existing methodologies. The conducted experiments are for limited number of iterations and with limited computational resources.

2. Multi-Parent Recombination Evolution Strategy

The MPR-ES is from the type $\mu/\rho+\lambda$ [6]. Where μ is the number of generated parent individuals in the generation G . λ is the number of children individuals in the generation G . And ρ is the mutation and/or the recombination strategy used to produce children from the parents. The $\mu/\rho+\lambda$ mean that μ parents ‘produce’ λ children using recombination and/or mutation. Each of the μ parents and λ children is then assigned a fitness value (the value of the objective function). Those with best fitness (the fittest) both parents and children become next generation parents. This is the so called *multimembered ES* proposed by *Rechenberg* and later elaborated by *Schwefel* [2][8].

A maximum number of generations, achieving required optimal value or complex condition could be used as a termination criterion.

To design of MPR-ES is based on the concept have to make the following assumption. Let $S_p(M)$ be the parent individuals array, $S_c(M)$ – the offspring (children) individuals array and G – the number of the generations. M is the number of dimensions (arguments) of the individuals. The pseudo code for the algorithms is as follows:

```

Begin
  Generations = 1
  Initialize (Parents)
  Evaluate (Fitness (Parents))
  While (! TerminalCondition) {
    Parents = ParentSelection;
    Children = Recombination (Parents);
    If (mutatuion_probability){
      Children=Mutation (Children)
      Evaluate (Fitness (Children))}
    Tournament Selection (Parents, Children)
    Generations++}
End

```

The algorithm starts its work by randomly generating S individuals (real numbers) for the parent array in pre-defined search space boundaries. After that the generation of offspring individuals (recombination) begins. What makes this MPR-ES distinctive is that the successors are generated by use of three solutions. This improves the probability to create stronger and more successful new solution. In other words – using this strategy is more likely the offspring to be “scattered” around the search space and more effectively to reach the optimum of the function. The three parent solutions are randomly chosen for each child from the parents array:

$$S_c(M) = S_{p\eta}(M) + S_{p\varepsilon}(M) + S_{p\phi}(M) / 3 \quad (1)$$

The given formula is for how the children’s array is generated where η , ε and ϕ are random numbers.

In this ESA the probability for mutation to occur is 60%. If this happens then the child individuals array is mutated. From the conducted experiments on different objective functions the most successful strategy happens to be the following one:

$$S_c(M) = S_c(M) * (X \min + random * (X \max - X \min)) \quad (2)$$

where $Xmin$ and $Xmax$ are the search space lower and upper boundaries.

Similar to nature’s mutation where sometimes it helps for the faster adaptation of the individuals to the habitat, this mutation is very important, because it has the power to bring the algorithm out of local optima traps. After the recombination and the mutation (if it happens to occur) follows the *tournament selection function* (the self-adaptation strategy for this MPR-ES). This process is of great importance, because it guarantees that only the fittest (best) S individuals both from the parent’s or the children’s array will survive and will be able to reproduce in the next generation. The way it works is simple. Based on the fitness functions assigned to each individual only the best S are selected. Each of these steps is repeated in a loop until the stop criterion is met. For our case when the maximum number of generations is reached. In order to assess MPR-ES it is compared to Real-Coded Genetic Algorithm Blend Crossover alpha modification [12] with variable blend alpha [7].

3. Real Coded Genetic Algorithm

A computational implementation and application of Genetic Algorithms is proposed by Holland [13]. Genetic algorithms are different from other optimisation and search processes in several ways:

- (1) GAs work with a coding of the parameter set, not the parameters themselves;
- (2) GAs search from a population of points, not from a single point;
- (3) GAs use payoff (objective function) information, not derivative or other auxiliary knowledge;
- (4) GAs use probabilistic transition rules, not deterministic rules [14].

The Genetic Algorithm begins with initialisation. Initialisation is a stochastic selection and evaluation of a set of initial solutions, called initial population. The next step is to form a new population. Generation of a new population consists of the events modification, replacement and evaluation.

During the modification, individuals from the current population are taken and used for creation of an offspring. Most often for generation of a new individual, two individuals from the current population are selected. These individuals are called parents. Recombination between selected parents produces an offspring. This is motivated by expectations, that the new individual can be better than its parents. To generate a solution different from selected parents the offspring can be a subject of mutation.

Mutation generates a random value for the offspring. Mutation can appear with certain, small probability. After modification follow the events evaluation and replacement. The Genetic Algorithm uses a preliminary defined objective (fitness) function included as part of the evaluation. For any particular optimisation problem the fitness function is different. After the event replacement, follows the event termination. Unless termination criterion (for example expiration of the number of iterations or improvement of the best solution) is met, repetition of the events modification and replacement continues.

Used in this study GA is with Blend crossover modification strategy called also BLX-alpha [12]. For BLX-alpha modification strategy, the offspring is a random location within the area determined by selected parents and extended with a blend interval alpha. The mathematical description of BLX-alpha modification strategy is presented with equation below.

$$X_c = X_{p1} - \alpha + (X_{p2} - X_{p1} + 2) * \text{random}(0,1) \quad (3)$$

where X_{p2} and X_{p1} are selected parents, $X_{p2} > X_{p1}$, α is a blend around the selected parents, $\text{random}(0,1)$ generates a random value between 0 and 1.

The extension of the space, between selected parents, increases the chances of the algorithm to reach the appropriate solution if it is near to the area determined by the parents. Increasing the chances to reach the optimum is an advantage. Variation of the blend α can be used for tuning of the convergence and divergence of the search process. Therefore, the concept for extension of the space for modification by a blend α is considered as valuable for improvement of the performance of the search process.

For the purposes of the investigation the Genetic Algorithms BLX- α is modified and implemented with a variable blend α [7].

4. Black Box Model for Test Definition

For this study in order to guarantee fair and equal condition for exploration or any search task a model for Black box definition of optimisation test is applied for both Multi-Parent

Recombination Evolution Strategy and Real Coded Blend Crossover Genetic Algorithm is applied.

The main characteristic of used Black box model is that the objective and constraint functions are explicitly not known for the search method [15].

Applying the Black box concept to numerical problems involves implementation of an interface between data space defined by the objective function and the optimisation process. If optimisation methods have no abilities for adaptation the black box may not be feasible. Both used methods are implemented in a manner, which guarantee sufficient for black box search level of adaptation. An ability of the algorithm to explore, to learn and to cope with unknown problems usually leads to the improvement of performance and effectiveness of the optimisation process.

Presented study uses four test functions for testing MPR-ES and comparing it to GA BLX alpha.

All tests are implemented in Black Box model and do not require modification of the methods or tuning their parameters for each particular test. MPR-ES and GA BLX alpha have to adapt to explored tasks.

5. Test Functions

The numerical test functions used for experiments are presented in this section; these are Ackley test [3], Rosenbrock test [9], Rastrigin test [11], and Griewank test [5].

The objective function of Ackley test [3] is:

$$f(x) = 20 \exp \left(0.2 \sqrt{\frac{1}{D} \sum_{i=1}^D x_i^2} \right) \exp \left(\frac{1}{D} \sum_{i=1}^D \cos(2\pi x_i) \right) + 20 + \exp(1) \quad (4)$$

The maximal value is $f_{max}(0.0, 0.0) = 0.0$.

The Rosenbrock test function [11] is:

$$f(x) = \sum_{i=1}^{D-1} \left(100(x_i^2 - x_{i+1})^2 + (x_i - 1)^2 \right) \quad (5)$$

The maximal value is $f_{max}(1.0, 1.0) = 0.0$.

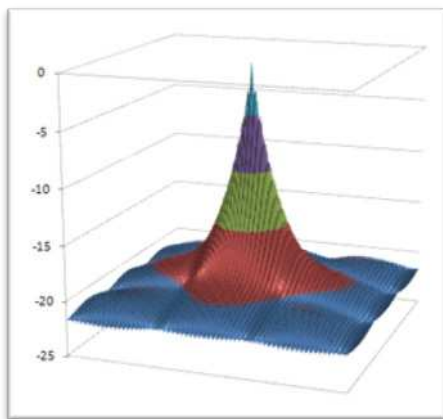


Figure 1: Ackley function

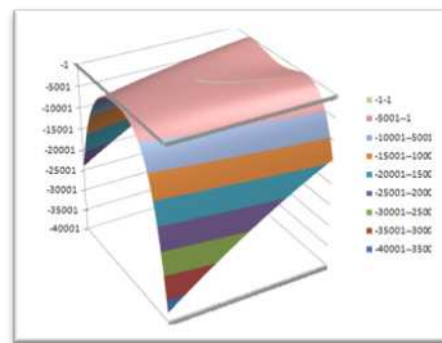


Figure 2: The Rosenbrock function

The objective function Rastigin test [9] is:

$$f(x) = \sum_{i=1}^D \left(x_i^2 - 10 \cos(2\pi x_i) \right) - 10 * D \quad (6)$$

The maximal value is $f_{max}(0.0, 0.0) = 0.0$.

The Griewank test objective function [5] is:

$$f(x) = \sum_{i=1}^D \frac{x_i^2}{4000} - \prod_{i=1}^D \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1 \quad (7)$$

The maximal value is $f_{max}(0.0, 0.0) = 0.0$.

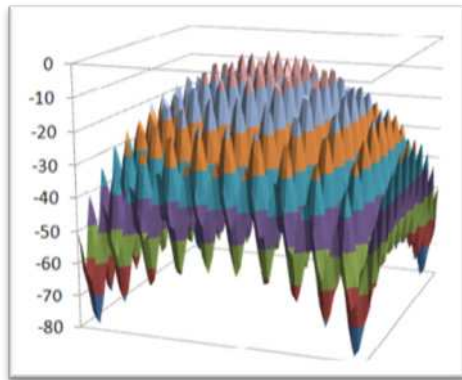


Figure 3: The Rastrigin function

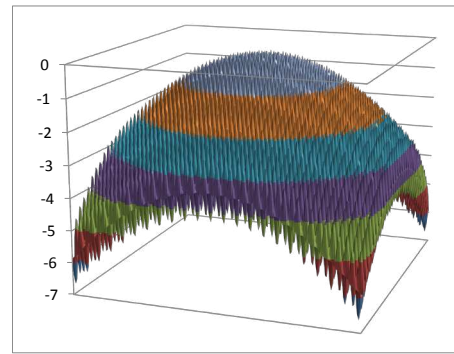


Figure 4: The Griewank function

6. Experimental Results

All experiments on the objective functions for optimization with Multi-Parent Recombination Evolution Strategy and Real Coded Genetic Algorithm Blend Crossover alpha with variable blend are conducted with the same initial parameters for both algorithms. The parameters are as follows:

- The number of individuals in a generation is 10. This is due to demonstrate that even a small number of random generated individuals is capable of reaching the optimum by recombination, mutation and adaptation.
- All tests are used in their two dimensional $M=2$ variants.
- The number of maximum generations for MPR ES limited to 2000 and for GA BLX alpha to 20000 accordingly. The reason for this is that for one iteration MPR ES produces 10 new individuals and GA BLX alpha only one. These numbers of iterations guarantee equal number of fitness function calculations required for fair comparison.
- Probability for mutation to occur for MPR ES is 60% and for GA BLX alpha 40%. The percentages are relatively high but since mutation is of a great importance for global optimization they are justified. Blend alpha for GA BLX varies from 0.5 to 1.5 with step 0.1. With variable blend GA BLX alpha produces 320 results per test. For comparison is used best achieved result.

Although the results are shown for the 2000-th generation in some cases the MPR ES leads to the presented result before the 500-th generation.

Experimental results are presented in the tables below.

Table 1: Ackley test		
Results for MPR ES		
Search range	Optimum	2000G Result
[-32; 32]	0	-0.000004
Results for GA BLX alpha		
[-32; 32]	0	-0.000202

Table 2: Rosenbrock test		
Results for MPR ES		
Search range	Optimum	2000G Result
[-1000; 1000]	0	-0.000000
Results for GA BLX alpha		
[-1000; 1000]	0	-0.000000

Table 3: Rastrigin test		
Results for MPR ES		
Search range	Optimum	2000G Result
[-1000; 1000]	0	-0.000000
Results for GA BLX alpha		
[-1000; 1000]	0	-0.000101

Table 4: Griewank test		
Results for MPR ES		
Search range	Optimum	2000G Result
[-1000; 1000]	0	-0.000000
Results for GA BLX alpha		
[-1000; 1000]	0	-0.040255

Experimental results confirm the expectations that Multi-Parent Recombination Evolution Strategy will adapt to explored numerical tests.

MPR ES resolves Ackley test, Rosenbrock test, Rastrigin test, and Griewank test without resetting or retuning of its optimisation parameters with precision above 0.0000001.

For comparison and Real Coded Genetic Algorithm Blend Crossover alpha resolves Rosenbrock test with the same precision, Ackley test and Rastrigin test with precision 0.0001, and for Griewank test needs additional iterations or perhaps retuning of mutation probability.

7. Conclusions

The Multi-Parent Recombination Evolution Strategy reviewed in this paper presents a distinctive approach in the offspring generation from three different parent individuals. The conducted experiments showed that this strategy, along with the mutation strategy successfully leads the algorithms to the optimal value of most of the objective functions even before it has reached the present maximum number of iterations (generations). Also equipped with the self-adaptation strategy and supported by it guarantees that each generation is at least as good as the previous one if not better.

The conducted experiments and the results from them show that there is no need of a large number of individuals in a generation to achieve the objective – in our case reaching the optimum.

Moreover the small number of individuals in the population combined with the relatively simple strategies for recombination, mutation and self-adaptation leads to very fast computation of the solutions.

This in turn opens possibilities before the MPR ES for use in dynamic real-time applications such as transmit power minimisation in wireless communications, in Quality of Service (QoS) related to IP telephony where specific requirements on signal-to-noise ratio and loudness levels are needed. Also it has great potential in the field of pattern recognition in broadcast signals (wireless networking).

The results from the experiments show that this Multi-Parent Recombination Evolution Strategy based on its simplicity can turn in a powerful tool for optimization.

Further research could focus on evaluation of MPR ES on multidimensional, constrained and time dependent tests.

It will be a challenge to see how Evolution Strategies perform on numerical tests where the optimum is unknown and to compare their performance to other search algorithms.

References

1. Beck, T., 1996, "*Evolutionary Algorithms in Theory and Practice: Evolution strategies, Evolutionary programming, genetic algorithms*", Oxford.
2. Beck, T., and Hoffmeister, F., and Schwefel, H.-P., 1991, "*A survey of evolution strategies*", University of Dortmund, Department of computer science.
3. D. H. Ackley., 1987 "*A connectionist machine for genetic hillclimbing*". Boston: Kluwer Academic Publishers.
4. Fogel, D., B., 2002, "*Evolutionary Computing*", IEEE Press.
5. Griewank, A. O., 1981 "*Generalized Decent for Global Optimization.*" *J. Opt. Th. Appl.*, Vol. 34, pp. 11-39.
6. Beyer H.G., Schwefel, H.P., Evolution Strategies: A Comprehensive Introduction, In: Natural Computing: an international journal, Vol. 1, No. 1 Hingham, MA, USA: Kluwer Academic Publishers, 2002, pp. 3-52.
7. Penev, K., Littlefair, G., 2003, "*Free search a novel heuristic method*", Solent university Southampton, pp. 133-134.
8. Rechenberg, I., 1973, "*Evolutionsstrategie: Optimierung technischer Systeme und Prinzipien der biologischen Evolution*", Frommann-Holzboog, Stuttgart.
9. Rosenbrock, H. H., 1960, "*An automatic method for finding the greatest or least value of a function*", *The Computer Journal*, Vol. 3, pp. 175-184.
10. Schwefel, H.-P., 1995, "*Evolution and Optimum Seeking*", Wiley.
11. Törn and A. Zilinskas., 1989 "*Global Optimization*". Lecture Notes in Computer Science, Vol. 350, Springer-Verlag, Berlin.
12. L. J. Eshelman, J. D. Schaffer, Real-coded genetic algorithms and interval-schemata, Foundations of Genetic Algorithms 2, Morgan Kaufman, pp 187-202, San Mateo, 1993.
13. Holland J., 1975, Adaptation in Natural and Artificial Systems, University of Michigan Press.
14. Goldberg D.E., 1989, Genetic Algorithms in Search, Optimisation, and Machine Learning, Addison Wesley Longman Inc. ISBN 0-201-15767-5.
15. Brekelmans R., Driessen L., Hamers H., Hertog D., 2001, A new Sequential Optimisation Approach to Product and Process Design Involving Expensive Simulations, Third ASMO UK / ISSMO Conference on Engineering Design Optimisation.

Real Coded Genetic Algorithm with Enhanced Abilities for Adaptation Applied to Optimisation of MIMO Systems

Gergana Yalamova¹, Georgi Iliev², Kalin Penev¹

¹Technology School, Southampton Solent University,

²Faculty of Telecommunications, Technical University of Sofia

¹gdyalamova@yahoo.com, ²gli@tu-sofia.bg, ³Kalin.Penev@solent.ac.uk

Abstract: This article presents an investigation of real coded Genetic Algorithm Blend Crossover Alpha modification, with enhanced ability for adaptation, applied to minimisation of transmit power in multiple-input multiple-output (MIMO) systems beamforming. The goal is to formulate transmit power minimisation task as a black box software object and evaluate an alternative to currently existing methods for optimisation of transmit energy in multicast system constrained by signal to noise ratio. The novelty of this adaptive methodology for determination of minimal power level within certain Quality of Service criteria is that it guarantees satisfaction of the constraint and 100% feasibility of achieved solutions. In addition this methodology excludes retuning algorithms parameters by using black box model for the problem definition. Experiments are conducted for identification of weight vectors assigned for signal strength and direction. Achieved experimental results are presented and analysed.

Keywords: MIMO Multicast Systems, Transmit Power Minimisation, Optimisation, Real Coded Genetic Algorithm BLX alpha.

1. Introduction

Due to the powerful performance and due the abilities to achieve a high level of spectral efficiency required in 3G and 4G technologies, MIMO systems are popular in wireless networks [3][13][18]. These systems become an integral part of Evolved High-Speed Packet Access (HSPA+), Worldwide Interoperability for Microwave Access (WiMAX) and Long Term Evolution (LTE) mobile technologies [17][18]. Continuous growth of modern multimedia applications and extension of high data rate transmission towards multiple users requires effective use of limited resources and reduction of power emission. In this case MIMO systems are particularly useful. The capacity and quality of wireless communications enhancement by using multiple-output antenna arrays configuration are widely discussed in the literature [6][10][11][19]. A simultaneous serving multiple mobile stations (users), without compromising available radio spectrum, by a base station (BS) in a MIMO system, needs to perform optimal beamforming within Quality of Service (QoS) constraints in order to suppress multiuser interference (MUI) to the end users [8]. Maximizing overall capacity in compliance with the utilisation of available Channel State Information (CSI) at the transmitter additionally improves the performance but complicates the energy efficiency problem formulation [2][9][12]. The requirements for minimal power, limited frequency bandwidth and hard signal to noise ratio constraint make in practice the task for designing high data rate of wireless communication systems extremely challenging [20]. The transmission power as a very substantial resource requires adequate antenna management methods that in addition have to satisfy the defined signal-to-interference-plus-noise ratio (SINR) constraints. There are many methods and strategies, which are searching for solutions of this optimisation problem by reducing the computational complexity, relaxation techniques based on semidefinite programming (SDP), interior point methods in polynomial time, minimal transmit power indicators, and multicast power control schemes [2][8].

The essence is that optimisation faces a dilemma between transmit antennas number increase and signal to noise ratio satisfaction. A higher number of transmit antennas refers to maximization of the system reliability, expansion of the coverage area and decrease of the required transmit power but it complicates the task for determination signal for each antenna.

Although many methods for resolving this problem are published, there is not enough evidence that they could provide a solution convenient for wide implementation in practice, which can manage with the hard SINR constraint. QoS (providing a guaranteed minimum received SINR to every user) and max-min fair (MMF) (maximizing the smallest received SINR) define that the transmit beamforming for multicast systems problem is NP-hard [16].

This investigation applies real coded Genetic Algorithm Blend Crossover Alpha modification (GA BLX α) [4] with enhanced abilities for adaptation (GA BLX α EAA) [14] to transmit beamforming power minimisation. The task core is formulation of the problem into software object using Black box model for task definition required by algorithms with enhanced abilities for adaptation, capable to cope with hard constrained non-linear numerical tests at this level of complexity [14]. Evaluation of the algorithms' abilities in terms of fast identification of acceptable minimal power, more accurate results, and adaptation to parameters variation, and possibilities for further implementation and integration within real MIMO systems are the main objective of this study. A number of experiments are conducted for identification of weight vectors assigned for signal strength and direction.

Within the next section the theoretical problem formulation and antenna configuration system model are introduced. A methodology for determination of minimal power level for certain QoS criteria is discussed. Compact overviews of Genetic Algorithm and Black box model for task formulation are also presented within the second section. The third section summarises the results in tabular and graphical format followed by critical analysis and discussion on essential characteristics of the used method. Recommendations for future research conclude the article.

2. System Model and Problem Statement

Notations: In the sections below, lowercase bold letters denote column vectors, in particular channel and weight vectors. The Hermitian transpose matrix operator is denoted as $(\cdot)^H$. The Euclidian norm is indicated as $\|\cdot\|_2$ while the absolute value is denoted as $|\cdot|$.

2.1. MIMO Antennas Configuration

This section presents a system model based on previous study [15] and beamforming management strategy assuring QoS. The model includes a single-cell MIMO mobile system with N antenna elements of a multi-antenna BS and M mobile stations (MS) with single antenna receivers (for each users). Regarding the fact that the system could perform transmission of multiple diverse data signals - $s_k(t)$, the users are sectioned in different multicast groups indicated below with G . Mobile stations that receive the same information defined a particular Group. The number of groups should be in the range: $1 \leq G \leq M$ which specifies three kinds of scenario: $G = 1$ broadcasting, $1 < G < M$ multicasting, $G = M$ unicasting.

The channel state information (CSI), contained by channel vectors $\mathbf{h}_i \in \mathbb{C}^N, \forall i \in \{1, \dots, M\}$ is used for beamforming at the transmitter. The channel is assumed to be frequency-flat quasi-static channel and the propagation loss and phase shift of it depends on these vectors as well as the quality of the CSI at the transmitter impacts beamforming performance.

Each pair of transmit-receive antennas provides a signal path from transmitter to receiver. By sending the same information through different paths, multiple independently-faded replicas of data can be obtained at the receiver end. Hence, more reliable reception is achieved [15].

$$\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_M = \begin{bmatrix} h_{11} \\ h_{21} \\ \vdots \\ h_{N1} \end{bmatrix}, \begin{bmatrix} h_{12} \\ h_{22} \\ \vdots \\ h_{N2} \end{bmatrix}, \dots, \begin{bmatrix} h_{1M} \\ h_{2M} \\ \vdots \\ h_{NM} \end{bmatrix}$$

The instantaneous signal that contains information for the users in group k at the moment t with noise variance σ_i^2 is $s_k(t)$. The adaptation of the energy strength and direction is achieved by selecting a number of appropriate weight vectors corresponding to each antenna element [1]. Multiplication of the signal with these weight vectors at the transmitter is necessary to confirm the required signal-to-interference-plus-noise ratio (SINR) levels, c_i , $\forall i \in \{1, \dots, M\}$. The weight vectors are denoted as $\mathbf{w}_k \in \mathbb{C}^N \forall k \in \{1, \dots, G\}$ then:

$$\sum_{k=1}^G \mathbf{w}_k^H s_k(t) \quad (1)$$

This formula describes that the transmitted signal by the BS is a linear combination of weight vectors and the signals for each group.

Assuming that $s_k(t)$ is zero-mean as well as that $\{s_k(t)\}_{k=1}^G$ are mutually uncorrelated, which means the total transmitted energy is:

$$\sum_{k=1}^G \|\mathbf{w}_k\|_2^2 \quad (2)$$

2.2. Problem Statement

The problem of minimizing the total transmitted power [1, 7] by covering the defined in advance SINR requirements for each user is presented as:

$$\begin{aligned} & \underset{\{\mathbf{w}_k \in \mathbb{C}^N\}_{k=1}^G}{\text{minimize}} && \sum_{k=1}^G \|\mathbf{w}_k\|_2^2 \\ & \text{subject to:} && \frac{|\mathbf{w}_k^H \mathbf{h}_i|^2}{\sum_{l \neq k} |\mathbf{w}_l^H \mathbf{h}_i|^2 + \sigma_i^2} \geq c_i, \\ & && \forall i \in G_k, \forall k, l \in \{1, \dots, G\}. \end{aligned} \quad (3)$$

Earlier publication [12] where a relaxed method based on an equivalent problem is applied, suggests that it is difficult to cope with the SINR constraint in a reasonable time. This investigation attempts to evaluate real-value coded GA BLX α EAA on transmitted power minimisation without task relaxation and with 100% satisfaction of SINR constraint.

2.3. Methodology

Methodology used for optimisation of MIMO systems, as it is mentioned above, is based on real coded Genetic Algorithm. According to the literature GAs are a family of computational models inspired by natural selection and evolution [7]. GAs, which operates on a set of solutions [5] are recognised as valuable in the domain of Computational Intelligence and in optimisation of hard tasks. This investigation focuses on BLX modification of real coded Genetic Algorithm [4] with variable blend crossover α [14].

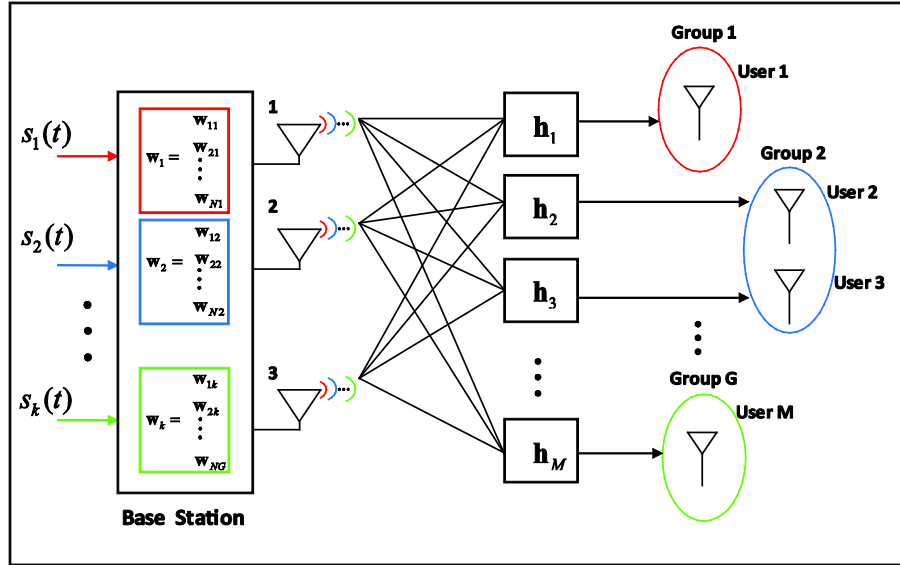


Figure 1: Diagram of transmit beamforming antenna configuration.

Configuration of the algorithms parameters including population size, blend variation, selection and replacement are rigorously tested on heterogeneous numerical tasks and together with the results are published [14]. In used Black box search model the objective function is differentiated and separated from the algorithm. Tasks are designed as replaceable software objects - black boxes independent to the method. This makes the methodology convenient for minimisation of transmit power where users change their locations and hard SINR constraint may changes its level over time.

The algorithm generates checks and delivers variables values to the problem 'box' within the assigned QoS constraints. Identification and usage for optimisation only variables, which satisfies SINR constraints guarantee that the achieved results are feasible.

The problem 'box' reacts and returns corresponding value of the objective function. In this case GA does not require knowledge about the explored tasks, including knowledge about possible hidden convexity. GA uses existing and newly generated values of the objective function for crossover and mutation in order to continue the optimisation process. Detailed description of the methodology is discussed earlier [14]. The search space borders and constraints are taken into account for generation of new variables. In fact experimental results indicate that once variables, which satisfy the SINR constraints, are identified then the optimisation is simple.

A previous study suggests that the convergence speed and overall success in GA depend on the initial population, which usually is stochastic [14]. Generation of this initial population could be related with already achieved solutions and with knowledge about the constraint restricted area if this knowledge exists. In case where the knowledge about feasible space of solutions is available, it supports generation of a better starting point for the algorithm and improves the GA's performance. Simply the algorithm does not generate candidate solutions, which do not satisfy presented constraints. The GA next stage modification is its major event that involves selection of parents, recombination between them and mutation. For selected Blend crossover modification strategy called BLX- α the offspring is a random location within the area determined by randomly selected parents and extended with a blend interval α . A mathematical description is presented with equation (4).

$$X_{\text{offspring}} = X_{p1} - \alpha + (X_{p2} - X_{p1} + 2\alpha) * \text{random}(0,1) \quad (4)$$

where X_{p2} and X_{p1} are selected parents, $X_{p2} > X_{p1}$, α is a blend around the selected parents, $random(0,1)$ generates a random value between 0 and 1 [14]. An extension of the space between selected parents increases the chances of the algorithm to reach an appropriate solution if it is near to the area determined by the parents. The population size for all experiments is 10 individuals where each individual corresponds to the set of weigh vectors for explored case. The blend α varies between 0.7 and 1.7. Offspring replaces the low fit individual of the population if it is better. Mutation probability is 0.3.

The design of MIMO transmit power objective function as a Black box software object allows adaptation to the explored problem and reduces necessity for preliminary settings of the optimisation parameters. This makes the methodology potentially useful for implementation in devices firmware and determines it selection for the investigation.

3. Results

According to the system model, the input data is as following: number of transmit antenna array $N = 5$, number of user groups $G = 3$, users in each group $G_k = 1$ and users in total $M = 3$. Initial channel characteristics are preliminarily defined as random values with a noise variance set to $\sigma_i^2 = 1$ for all channels. The SINR levels in the range of 6 to 12 dB with step 2 dB for all users are evaluated.

For each SINR level, series of experiments limited to 100, 500 and 1000 iterations are accomplished. Generated results are presented in Table 1. The first stage identifies and assesses convergence speed. The second stage evaluates the method reliability on minimisation of MIMO transmit power.

The convergence speed evaluation is based on four experiments with different level of SINR (6dB, 8dB, 10dB, and 12dB) where the number of tests is 10 for each experiment. These 10 tests differ from each other by the initial conditions namely starting from purposefully different stochastically generated initial populations.

The first experiment for lowest examined SINR level 6dB begins with randomly generated combination of weight vectors according to the equation:

$$\mathbf{w}_k(c_r) = \mathbf{w}_{\min} + \text{random}(0,1) * (\mathbf{w}_{\max} - \mathbf{w}_{\min}) \quad (5)$$

for $\forall k \in \{1, \dots, G\}$ and $\forall r \in \{1, \dots, R\}$ where weight vectors \mathbf{w}_k are in range between the lowest \mathbf{w}_{\min} and the highest \mathbf{w}_{\max} limit, which in this investigation is $[-2; 2]$. A number of different SINR levels limited to $R = 4$ are denoted with c_r . The algorithm searches for a combination, which produces minimal, power consumption.

In order to avoid generation of non-feasible candidate solutions, which do not satisfies presented constraint, the second experiment for SINR level 8dB starts from a slightly randomised set of weight vectors achieved for 6dB SINR level, which satisfies this constraint, according to the equation:

$$\mathbf{w}_k(c_{r+1}) = \mathbf{w}_k(c_r) + 0.0001 * \text{random}(0,1) \quad (6)$$

for $\forall k \in \{1, \dots, G\}$ and $\forall r \in \{1, \dots, R\}$. Utilisation of the knowledge about constrained and non-constrained area in generation of initial population reduces and avoids exploration of unfeasible candidates' solutions, which decreases the overall optimisation period of time. The algorithm searches, first, for combination of variables with satisfies the higher SINR constraint level and then identifies, which of them produces minimal, power consumption. The same process is repeated for the rest of the SINR constraints 10dB and 12dB respectively. This strategy is expired by real life vocal communications. Applied to MIMO

systems it reflects positively on the period of time for identification of these sets of parameters, which satisfies the SINR constraint.

The graphs for four experimental processes of the first stage are illustrated on Figures 2-5:

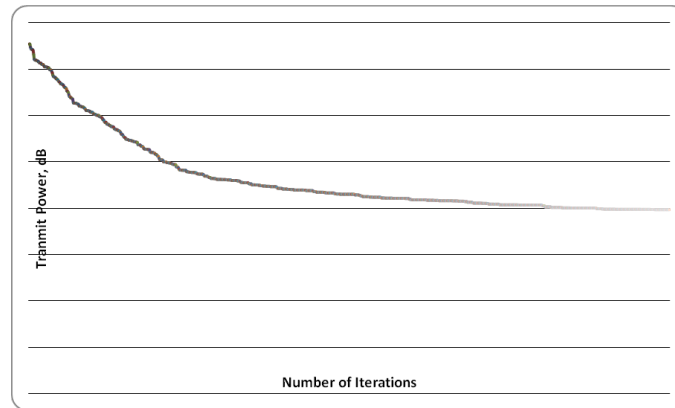


Figure 2: Achieved minimal Transmit Power for SINR level 6 dB

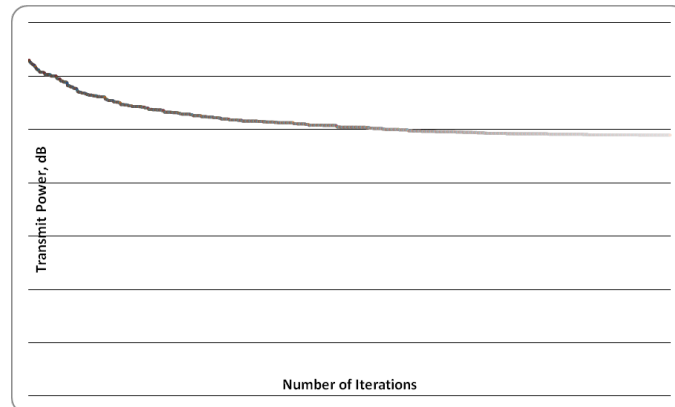


Figure 3: Achieved minimal Transmit Power for SINR level 8 dB

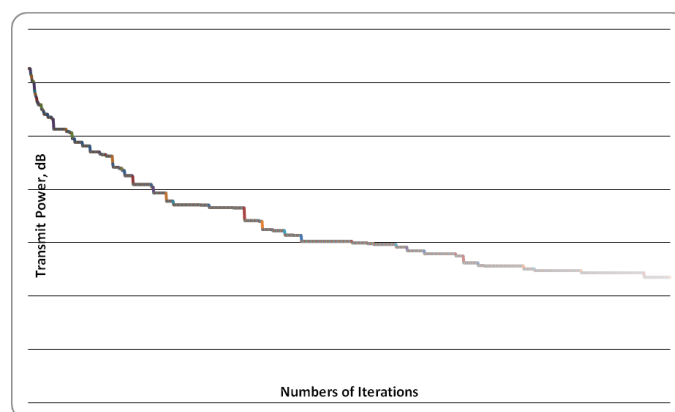


Figure 4: Achieved minimal Transmit Power for SINR level 10 dB

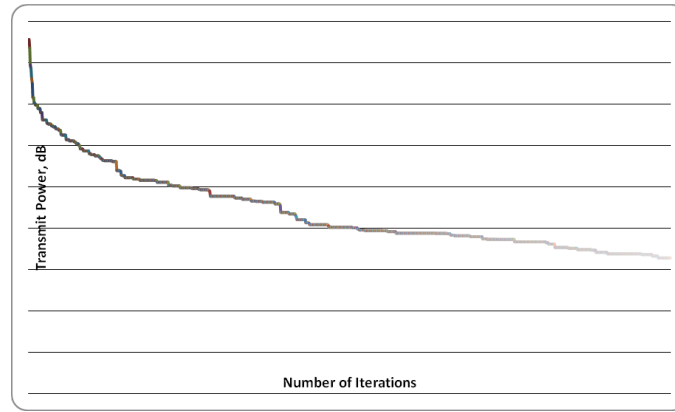


Figure 5: Achieved minimal Transmit Power for SINR level 12 dB

Presented above graphics indicates that GA BLX α EAA achieves acceptable level of transmit power within 500 iterations. Within 1000 iterations minimisation of the emitted power continues and reaches level achieved by other methods [15]. The advantage is that all the GA BLX α EAA results are 100% feasible.

Additional set of experiments with gradual change of the SINR levels starting from 0 dB to 12 dB with step 1 dB has been performed. In this case the process of power optimisation starts from stochastic initial population, which satisfies the constraints for 0 dB SINR level and uses the achieved results to generate feasible initial candidate solutions for next levels of the SINR constraint (6 dB, 8 dB, 10 dB, 12 dB). In order to evaluate probability for success the number of experiments is 320 for 10 different values of blend Alpha in total 3200 for each constraint level.

Achieved experimental results are presented in Table 1.

Table 1: SINR, dB	Iterations Number	Mean	Std	P_{\min} dB	P_{\max} dB	Feasibility, %
6	100	1.01	0.02	6.4646	6.5649	100
6	500	1.01	0.02	6.4269	6.5218	100
6	1000	1.01	0.01	6.4212	6.5013	100
8	100	1.02	0.04	8.6458	8.8007	100
8	500	1.01	0.03	8.5865	8.6802	100
8	1000	1.01	0.02	8.5724	8.6493	100
10	100	1.01	0.03	10.6761	10.7356	100
10	500	1.00	0.02	10.6754	10.7161	100
10	1000	1.00	0.02	10.6722	10.7100	100
12	100	1.00	0.05	12.7673	12.8313	100
12	500	1.00	0.03	12.7562	12.7920	100
12	1000	1.00	0.02	12.7493	12.7927	100

In Table 1, the first column presents the examined SINR levels in dB; the second column presents the number of iterations limit. In the third column, mean value refers to a normalised average value of the transmit power achieved for a particular number of iterations throughout all the experiments. This value is calculated from the sum of achieved power values after each experiment, divided by the number of experiments and normalised to the minimum obtained power P_{\min} . The standard deviation is denoted as *Std* in column 4.

The minimum power P_{\min} and the maximum power P_{\max} achieved for each case are shown in the column 5 and column 6 respectively. A solution is feasible when the achieved set of

variables satisfies the QoS constraints. Feasibility percentage of the successful solutions is indicated in column 7.

4. Discussion

Analysis of the experimental results suggests several issues, which deserve attention. The period of time required for identification of variables, which satisfies the SINR constraint for values above 6 dB is lower when the optimisation process is split on several sub-processes. It begins from low level of SINR (lower than 6 dB) and then gradually changes this SINR level using sets of parameters available from previous low level SINR. In this case the process performance seems to be less time consuming and allows acceleration of the overall optimisation.

A search process for SINR constraint level 6 dB and 8 dB limited to 100 iterations is completed for less than second and for 1000 iterations within 12 seconds. For SINR level 10 dB and 12 dB limited to 1000 iterations search process is completed for less than 3 seconds and for 1000 iterations within 22 and 60 seconds respectively. In case of rough change of SINR levels time for identification of variables, which satisfies this change increases. The gradual SINR change facilitates faster performance.

Precise assessment of periods of time required for constraint satisfaction and then for actual optimisation time, including comparison to other methods, could be a subject of further research. It could focus, also, on utilisation of more narrow methods for constraints satisfaction, in order to minimise the time for identification of variables, which satisfies the SINR constraint and lead to feasible solutions.

5. Conclusion

Presented in the article approach for using real coded Genetic Algorithm Blend Crossover Alpha modification with enhanced abilities for adaptation contributes to the research efforts in minimisation of transmit power for wireless MIMO systems. An overview of MIMO systems states the essence of beamforming transmit power minimisation. Transmit power minimisation task is implemented in Black box model suitable for optimisation by methods with enhanced abilities for adaptation. Particularly explored methodology demonstrates good capabilities to optimise transmit power satisfying 100% SINR hard constraint. The experimental results and their analysis suggest that in certain extent Genetic Algorithm could be an alternative for existing methods. Potential for improvement, in terms of decreasing the time for constraint satisfaction, is identified. For further research, the MIMO Black box software object could be developed as unified object for arbitrary number of antennas, users and users groups. Future investigation could focus, also, on comparison with other methods.

Acknowledgement

Preparation of this paper is supported by Southampton Solent University, Research and Enterprise Fund, Grants 516/17062011, 623/19102011.

References

1. J. Andrews, A. Ghosh, R. Muhamed, Fundamentals of WiMAX: understanding broadband wireless networking. Prentice Hall, 2007.
2. M. Bengtsson, B. Ottersten. Optimal and suboptimal transmit beamforming. Handbook of antennas in Wireless Communications. CRC Press LLC, 2002.
3. E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj and H. V. Poor, MIMO Wireless Communication. Cambridge University Press, New York, 2007.
4. L. J. Eshelman, J. D. Schaffer, Real-coded genetic algorithms and interval-schemata, Foundations of Genetic Algorithms 2, Morgan Kaufman, pp 187-202, San Mateo, 1993.

5. D. E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison Wesley Longman Inc. ISBN 0-201-15767-5, 1989.
6. A. He, S. Srikanteswara, K. Bae, T. Newman, J. Reed, W. Tranter, M. Sajadieh, M. Verhelst, Power Consumption Minimization for MIMO Systems – A Cognitive Radio Approach. IEEE Journal on Selected Areas in Communications, Vol. 29, No. 1, pp. 469-479, Jan. 2011.
7. J. Holland, Adaptation in Natural and Artificial Systems. University of Michigan Press, 1975.
8. E. Karipidis, N. D. Sidiropoulos, Z.-Q. Luo, Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups. IEEE Transactions on Signal Processing, vol. 56, no. 3, pp. 1268-1279, Mar. 2008.
9. B. G. Lee, D. Park, H. Seo, Wireless Communications Resource Management. John Wiley & Sons Pte Ltd, Singapore, 2009.
10. Q. Li, Q., G. Li, W. Lee, M. Lee, D. Mazzarese, B. Clerckx, Z. Li, MIMO Techniques in WiMAX and LTE: A Feature Overview. IEEE Commun. Mag., Vol. 48, No. 5, pp. 86-92, May. 2010.
11. A. Liu, H. Xiang, W. Luo, Y. Liu, Power minimization of multi-access MIMO systems with the rate constraint and finite – rate feedback. Information Sciences and Systems. CISS 43rd Annual Conference, Mar. 2009.
12. Z.-Q. Luo, E. Karipidis, N. D. Sidiropoulos, Transmit Beamforming to multiple co-channel multicast groups. IEEE International Workshop, 1:109-112, Dec. 2005.
13. C. Oestges, and B. Clerckx, MIMO Wireless Communications – from real-world propagation to space-time code design. Academic Press, Elsevier, Oxford, 2007.
14. K. Penev, Adaptive Search Heuristics Applied to Numerical Optimisation, PhD Thesis, Nottingham Trent University & Southampton Institute, UK, 2005.
15. N. Petrov, V. Poulkov, G. Iliev, Reduced Computational Complexity Beam-space Beamforming Design for MIMO Multicast Systems. NTMS, Cairo, Egypt, IEEE Dec.20-23, 2009.
16. N. D. Sidiropoulos, T. N. Davidson, Broadcasting with channel state information, in Proc. IEEE Sensor Array and Multichannel (SAM) Signal Process Workshop, Vol. 1, pp. 489-493, Sitges, Barcelona, Spain, Jul. 18-21, 2004.
17. W. Vereecken, W. V. Hendeghem, M. Deruyck, B. Puype, B. Lannoo, W. Joseph, D. Colle, L. Martens, and P. Demeester, Power Consumption in Telecommunication Networks: Overview and Reduction Strategies. IEEE Commun. Mag., Vol. 49, No. 6, pp. 62-69, Jun. 2011.
18. C.-X. Wang, X. Hong, X. Ge, X. Cheng, G. Zhang, J. S. Thompson, Cooperative MIMO Channel Models: A Survey. IEEE Commun. Mag., Vol. 48, No. 2, pp. 80-87, Feb. 2010.
19. Y.-H. Yang, S.-C. Lin, H.-J. Su, Multiuser MIMO Downlink Beamforming Design Based on Group Maximum SINR Filtering. IEEE Transactions on Signal Processing, Vol. 59, No. 4 pp. 1746-1758, Apr. 2011.
20. W. Yu, T. Lan, Transmitter optimization for the multi-antenna downlink with per-antenna power constraints. IEEE Transactions on Signal Processing, Vol. 55, No.6, Jun. 2007.

Overview of Routing Algorithms for MCP with Single or Mixed Metric

Yavor Tomov

Technical University of Sofia, Sofia, Bulgaria

qvor_tomov@abv.bg

Abstract: Nowadays more and more telecommunication services appear with increasing number of users. A large number of these services require a guaranteed Quality of Service (QoS). The main task of QoS routing is to find a path from source to destination satisfying all constraints. Our goal in this article is to make an overview of the algorithms able to solve Multi Constraint Problem (MCP) known as algorithms with single or mixed metrics.

Keywords: QoS, Routing Algorithm, MCP.

1. Introduction

As we all know the modern Internet from the very beginning was created on the principle of “The Best Effort”. This principle provides a fair usage of the net’s resources, but it is not able to cope with the new challenges, related to the provision of given Quality of Service (QoS).

According to the ITU definition the QoS is “The collective effect of service performance which determines the degree of satisfaction of a user of the service”.

Nowadays the number of services (especially such as the real time services) is constantly increasing. They need a guaranteed quality, which may be achieved by supporting one or more given parameters from source to destination.

The main tasks of routing are to find this path and to update and store the available data for a network.

A certain number of algorithms that are able to find the shortest path from a source to a destination are suggested in the literature [5][2]. These algorithms however can be applied only in cases when the parameter examined is one and only. When a certain service requires two or more parameters guaranteed from a source to a destination, the problem is known as Multi Constraint Problem (MCP). MCP is NP hard [7].

Depending on these parameters (such as delay, bandwidth, packet loss etc.) they may be classified as additive, multiplicative or concave. Delay is exemplified by an additive parameter, while the bandwidth is an example for concave. This paper focuses on the case with additive parameters, where the weight of this parameter (end to end) is equal to the sum of the weights of all links on the path. A path which satisfies all constraints is called a feasible path.

The algorithms, which are able to find this path (if it exists) are exact algorithms. The general drawback is their high complexity, and they are inapplicable in practice. They are mainly used for evaluating other types of algorithms known as heuristic. The heuristic algorithms refer to experience-based techniques for problem solving, they have lower complexity but they do not guarantee the finding of solution. The other class is ϵ -approximation algorithms. They are not necessarily exact but can provide a solution quantifiably close to the exact solution.

The article is aimed at reviewing a large number of algorithms, known as algorithms with single or mixed metrics. Most of them are heuristic but there are also exact ones. Their advantage is the low complexity and easy implementation in practice. Their drawback is the insufficient information they give concerning the fulfilment of QoS requirements [17].

2. Problem Formulation

Each network can be presented as a graph $G(V, E)$, where V is the set of nodes and E is the set of links. The nodes are routers or switches, while the links are physical or logical connections between them. Each link e is associated with n -dimensional link vector $\vec{w}(w_1, w_2, \dots, w_n)$. The source node will be noted as s , while the destination node will be noted as t . We will note the path from the source to the destination with p and the given constraints as L_i ($1 \leq i \leq n$). The problem is to find a path from source node to destination node such that

$$w_i(p) = \sum_{e \in p} w_i(e) \leq L_i \quad (1)$$

A path that satisfies all n constraints is referred to as a *feasible* path. Each path that has two metrics may be represented on the $((w_1(p), w_2(p)))$ plane.

Figure 1 shows that $w_1(p_2) < w_1(p_1)$ and $w_2(p_1) < w_2(p_2)$ and we cannot define whether $p_1 < p_2$ or not.

Figure 2 shows that $w_1(p_1) < w_1(p_2)$ and $w_2(p_1) < w_2(p_2)$. In this case the path p_2 is dominated by p_1 . If $w_i(p_1) < w_i(p_2)$ for all i , p_1 is the *dominant* path [5]. If there is a dominant path, it is easy to solve the MCP since having found the shortest path with respect to one parameter, the same path is the shortest with respect to all other parameters.

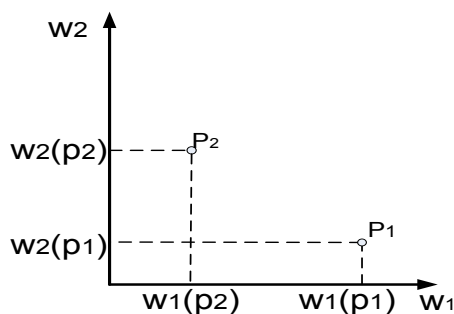


Figure 1: Non- dominant path

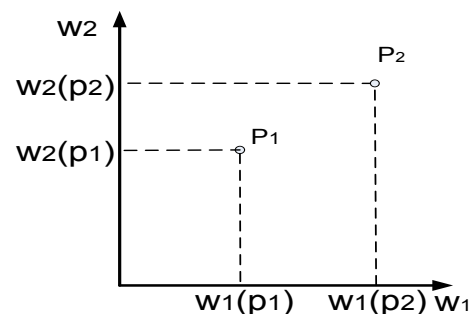


Figure 2: Dominant path

The algorithms that will be considered in this article use two different path lengths – linear path length (ALPL) and non-linear path length (ANLPL).

3. Algorithms with Linear Path Length.

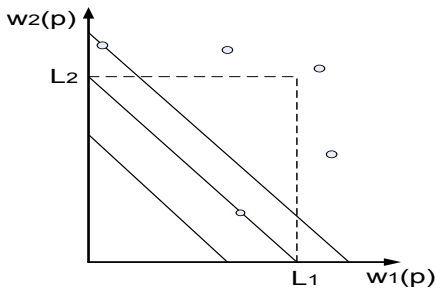
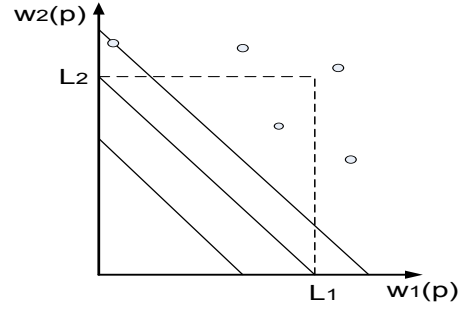
3.1. Jaffe's Algorithm

In 1985 Jaffe presented his heuristic algorithm [10], based on Lagrange relaxation. He suggested that the weights of each link be presented as linear combination.

$$w(e) = d_1 w_1(e) + d_2 w_2(e) \quad (2)$$

where d_1 and d_2 are positive multipliers.

The region between L_1 and L_2 (Fig. 3 and Fig. 4) is the feasible region of solutions.


Figure 3: The algorithm works

Figure 4: The algorithm fails

The black dots present the different paths. All paths that lie on the same parallel line have the same constant value c .

$$d_1 w_1(p) + d_2 w_2(p) = c \quad (3)$$

Each parallel line that lies above the other line has a higher constant value c . The main task in this method is defining the multipliers d_1 and d_2 . Jaffe proposed these multipliers to be defined by the following equation:

$$\sqrt{\frac{L_1}{L_2}} = \frac{d_2}{d_1} \quad (4)$$

Figure 3 indicates that Jaffe's algorithm works, while in Figure 4 it fails.

3.2. Feng's Algorithm

Based on Jaffe's heuristic algorithm Feng proposed a new exact algorithm [6], applying in it two basic ideas:

- Reducing the search space;
- k-th shortest path.

This algorithm reduces the search space as follows: firstly it finds the shortest path with respect to each weight, and the feasible region is defined as $L_i - w_i(p)$, Figure 5.

Secondly the author suggests that the multipliers be defined as:

$$d_i = \begin{cases} \infty & : L_i = w_i(p_i) \\ \frac{L_0 - w_0(p_0)}{L_i - w_i(p_i)} & : \text{otherwise} \end{cases} \quad (5)$$

The algorithm uses k-th shortest path, which means that if the first path returned by the algorithm is out of the feasible region it returns the second one and so on until the returned path by the algorithm is feasible or this path lies on a straight line with a higher constant value than the straight line with optimal constant value,

$$w_0(e) + \sum_{i=1}^{n-1} d_i w_i(e) > L_0 + \sum_{i=1}^{n-1} d_i L_i \quad (6)$$

then the algorithm returns no solution.

3.3. Iwata's Algorithm

Iwata proposed a heuristic algorithm (8) using single metrics to solve MCP. This algorithm uses Dijkstra's algorithm. It first computes one shortest path, based on a QoS measure and checks whether the constraints are satisfied. If they are satisfied – the algorithm returns this path. If they are not – another measure is applied and this is repeated until a feasible path is found. Figure 6 show that algorithm fails.

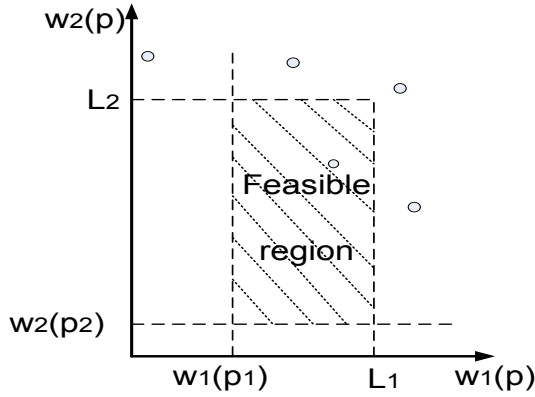


Figure 5: Reducing search space

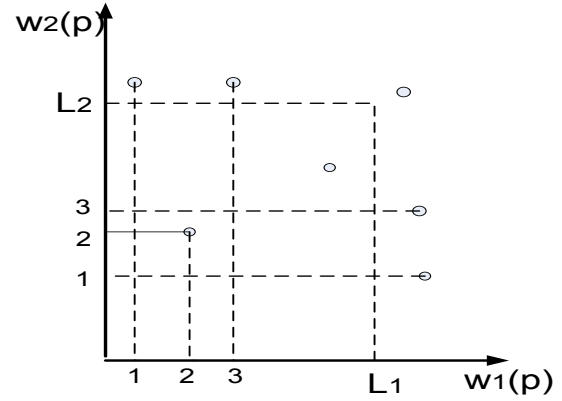


Figure 6: The algorithm fails

3.4. LARAC Algorithm

LARAC (Lagrange Relaxation based Aggregate cost) is a heuristic algorithm that is able to solve DCLC (Delay Constrained Least Cost) problem in polynomial time [11].

The first step in this method is to calculate the shortest path with respect to the cost- w_1 Fig. 7. In our case this path is p_2 and if this path meets the delay constraint $w_2(p_2) < L_d$, where L_d is constraint, this path is optimal. If this path doesn't satisfy this requirement, in the second step the algorithm finds the path with respect to delay constraint p_1 (if such a path exists). Thus the authors reduce the search space Fig. 7.

Based on the reduced search space the authors propose the multiplier to be defined in the following way:

$$d = \frac{w_1(p_2) - w_1(p_1)}{w_2(p_1) - w_2(p_2)}, \quad (7)$$

where w_1 , w_2 are the cost and delay respectively.

3.5. Khadivi's Algorithm

Khadivi proposed a heuristic algorithm [12], where he uses the single mixed metric as follow:

$$w(e) = \mu(e)[\Delta(e) + \varepsilon], \quad (8)$$

where ε is constant $0 \leq \varepsilon \leq 1$,

$$\mu(e) = \frac{1}{n} \sum_{i=1}^n \frac{w_i(e)}{L_i}, \quad (9)$$

and

$$\Delta(e) = \sum_{i=1}^n \left(\frac{w_i(e)}{L_i} - \mu(e) \right)^2, \quad (10)$$

where $w(e)$ is a new mixed metric.

4. Algorithms with Non-linear Path Length

4.1. TAMCRA (Tunable Accuracy Multiple Constraints Routing Algorithm)

Hans de Neve and Piet van Mieghem created a heuristic algorithm [4], based on three concepts:

- non-linear path length[1];
- k-th shortest path algorithm[3];
- non-dominated paths.

The authors replace the straight scanning lines used in the linear algorithms by curved equivalents lines Fig. 8.

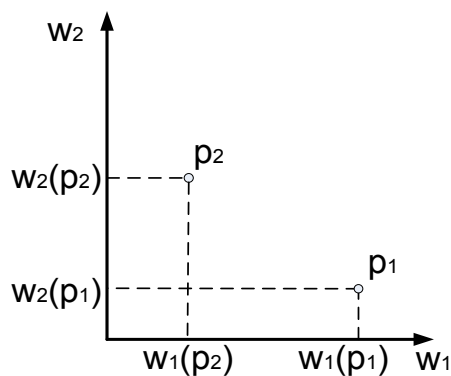


Figure 7: Two shortest paths w.r.t. cost and delay

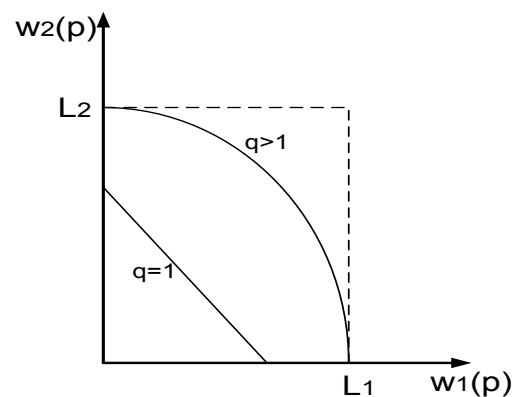


Figure 8: Linear and non-linear search space

In linear algorithms if the multipliers are calculated by the equation

$$\frac{L_1}{L_2} = \frac{d_2}{d_1} \quad (11)$$

then the area that will be scanned before the algorithm is able to return any solution outside the feasible region, will be half of the whole area.

The path's length from source to destination is non-linear combination:

$$l(p) = \sum_{i=1..n} \left(\frac{w_i(p)}{L_i} \right)^q \quad (12)$$

The best match is obtained in the limit when $q \rightarrow \infty$.

If we use the non-linear definition of the path length, *the subsection of shortest paths is not necessarily the shortest paths*. In this reason if we use Dijkstra's algorithm, it can fail to find solution. TAMCRA avoids this drawback using the k-th shortest path. The third feature of TAMCRA is that it stores the paths only if they are not dominated by the others.

4.2. H_MCOP

Kormaz and Krunz proposed a heuristic algorithm based on three theorems that they proved[13]. The algorithm uses two modified versions of Dijkstra's algorithm – in the forward direction and in the backward direction. In backward direction the algorithm uses the linear path length ($q = 1$) with respect to

$$w(e) = \sum_{i=1..m} \left(\frac{w_i(p)}{L_i} \right)^q \quad (13)$$

from each intermediate node to destination node t . After that the algorithm evaluates the suitable paths. In the forward direction the algorithm discovers each intermediate node from the source node, using the non-linear function ($q > 1$). The path from s to t in the forward direction is heuristically found.

4.3. SAMCRA

SAMCRA (Self-Adaptive Multiple Constraints Routing Algorithm) is a further development of TAMCRA, it is improved and exact algorithm [16]. Two versions were created. According to the first version the basic difference with TAMCRA is the fact that TAMCRA stores an equal number of paths at each node that are predefined, while SAMCRA adaptively defines the number of the paths at each node. Predefining the number of paths is TAMCRA main drawback because if this number is not sufficient then the algorithm fails to give solution.

In the second version the concept *look ahead* was added to reduce the search space. This concept was primarily proposed in [14][15].

5. Conclusions

The algorithms analyzed in this article are ALPL or ANLPL.

The major advantage of ANLPL is that when $q \rightarrow \infty$, they scan the feasible region precisely. Their drawback is that *subsections of shortest paths are not necessarily shortest paths*. Due to this, using Dijkstra's algorithm ANLPL are likely to fail. That is why algorithms using non-linear path length (TAMCRA, SAMCRA), need a modified Dijkstra's algorithm that stores k-number of paths. All this results in a higher algorithm complexity.

Korkmaz and Krunz[13] avoid the k-th shortest path concept. Instead they apply the "*look ahead*" concept to obtain paths which could be a possible solution. The usage of k-th shortest path guarantee that a shortest path will be found (if it exists) while "*look ahead*" does not.

When using ALPL only half of the feasible region is scanned before the algorithm is able to return the solution outside this region.

The main task is to choose the multipliers that define the ratio between the metrics, inside the mixed metric.

Iwata's algorithm [9] computes the shortest path with respect to each metric without mixing them. This algorithm has low complexity but the probability to find a solution is low too.

The authors of LARAC [11] use a similar method to find the shortest path with respect to cost and delay reducing the search space and based on this space to define a multiplier for a subsequent search with mixed metric.

Feng [6] applies the same approach, adding the concept of k-th shortest path to create an exact algorithm.

References

1. Ahuja R. K., Magnanti T. L., Orlin J. B., 1993, Network Flows: Theory, Algorithms, and Applications. Prentice Hall Inc.
2. Bellman R., 1958, On a Routing Problem. Quarterly Applied Mathematics, XVI(1), pp. 87-90.
3. Chong E. I., Maddila S., Morley S., 1995, On Finding Single-source Single-destination k-shortest Paths. Journal of Computing and Information, special issue ICCI' 95, pp. 40-47.
4. De Neve H., Van Mieghem P., 2000, TAMCRA: A Tunable Accuracy Multiple Constraints Routing Algorithm. Computer Communications, Vol. 23, pp. 66-679.
5. Dijkstra E. W., 1959, A Note on Two Problems in Connexion with Graphs. Numerische Mathematik, Vol. 1, pp. 269-271.
6. Feng G., 2003, Exact Algorithms For Multi-Constrained QoS Routing, Int. Conf. on Computer, Communication and Control Technologies (CCCT'03), Orlando, USA, Vol. 2, pp. 340-345.
7. Garey M. R., Johnson D. S., 1979, Computers and Intractability: A guide to the Theory of NP-Completeness. ISBN 0-7167-1044-7. W. H. Freeman and Company, San Francisco.
8. Henig M. I., 1985, The Shortest Path Problem with Two Objective Functions. European Journal of Operational Research, Vol. 25, pp. 281-291.
9. Iwata A., Izmailov R., Lee D., Sengupta B., Ramamurthy G., Suzuki H., 1996, ATM Routing Algorithms with Multiple QoS Requirements for Multimedia Internetworking. IEICE Trans. Commun., E79 - B(8), pp. 999-1007.
10. Jaffe J. M., 1984, Algorithms for Finding Paths with Multiple Constraints. Networks, Vol. 14, pp. 95-116.
11. Juttner A., Szviatovszki B., Mecs I., Rajko Z., 2001, Lagrange Relaxation based Method for the QoS Routing Problem, Proc. of the INFOCOM 2001 Conference, IEEE, Vol. 2, pp. 859-868.
12. Khadivi P., Samavi S., Todd T. D., 2008, Multi-constraint QoS Routing using a New Single Mixed Metrics, Journal of Network and Computer Applications, Vol. 31(4), pp. 656-676.
13. Korkmaz T., Krunz M., 2001, Multi-constrained Optimal Path Selection, Proc. of the INFOCOM 2001 Conference, IEEE, Anchorage, Alaska, Vol. 2, pp. 834-843.
14. Lin S., 1965, Computer Solutions of the Traveling Salesman Problem, Bell. Syst. Tech. J., 44(10), 2245-2269.
15. Newell A., Ernst G., 1965, The Search for Generality, in Proc. IFIP Congr., 1, 17-24, 1965.
16. Van Mieghem P., F. A. Kuipers, 2004, Concepts of Exact Quality of Service Algorithms, IEEE/ACM Trans. Net, 12, 851-862.
17. Wang Z., Crowcroft J., 1996, Quality of Service Routing for Supporting Multimedia Applications. IEEE Journal on Selected Areas in Communications, 14(7), 1228-1234.

Mono-objective Algorithm for Wireless Sensor Layout

Stefka Fidanova¹, Miroslav Shindarov², Pencho Marinov³
Institute of Information and Communication Technologies,
Bulgarian Academy of Sciences, Sofia, Bulgaria

¹stefka@parallel.bas.bg, ²miroslavberberov@abv.bg, ³pencho@parallel.bas.bg

Abstract: A wireless sensor network (WSN) is formed by sensors which communicate wirelessly. Each sensor node monitors area around itself and transmits collected data to a central gateway (high energy communication node), which provides a connection to the wired world where they can be collected, processed, analysed, and presented. In this paper we address a WSN layout problem instance in which full coverage is treated as a constraint, while objective function is a combination of the weighted number of the sensors and energy efficiency. To solve the problem we propose Ant Colony Optimization (ACO) algorithm.

Keywords: Wireless Sensor Network, Ant Algorithm, Mono-Objective Optimization.

1. Introduction

The sensors are small devices which can collect and transmit data. When the sensors communicate in a wireless manner, they create Wireless Sensor Network (WSN). Their working is based on the exchange of local information between nodes in order to achieve a global goal. WSN find applications in a variety of areas such as climate monitoring, military use, industry and sensing information from inhospitable locations [4][15]. For utilities such as the electricity grid, streetlights, and water municipalities, wireless sensors offer a lower-cost method for collecting system health data to reduce energy usage and better manage resources. In structural health monitoring, wireless sensors are used to effectively monitor highways, bridges, and tunnels. These systems can be deployed to continually monitor office buildings, hospitals, airports, factories, power plants, or production facilities. In a common WSN architecture, the measurement nodes are deployed to acquire measurements such as temperature, voltage, or even dissolved oxygen. WSN have so far been employed in military activities such as reconnaissance, surveillance, and target acquisition [5], environmental activities such as forest fire prevention, geophysical activities such as volcano eruptions study [16] and biomedical purposes such as health data monitoring [19] or civil engineering [12].

A WSN node contains components as radio, battery, microcontroller, analogue circuit, and sensor interface. In battery-powered systems, higher data rates and more frequent radio use consume more power. There are several open issues for sensor networks such as signal processing [12], deployment [18], operating cost, and localization and location estimation. All sensors must be able to transmit their data to HECN, either directly or via hops, using nearby sensors as communication relays. When deploying a WSN, the positioning of the sensor nodes becomes one of the major concerns. The coverage obtained with the network and the economic cost of the network depends directly of it. The task of selecting the geographical positions of the nodes for an optimally designed network can be very complex. Therefore, metaheuristics seem an interesting option to solve this problem.

In this paper we propose a solution method for the WSN layout problem using ACO. We focus on both minimizing the energy depletion of the nodes in the network and minimizing the number of the nodes, while the full coverage of the network and connectivity are considered as constraints. The problem is multi-objective. We convert it to mono-objective. The new objective function is a combination of the two objective functions of the original problem. The objective functions are weighted by parameter λ . In this paper we learn the influence of the parameter λ on the quality of the achieved solutions.

In [13] is solved an instance of WSN layout using a multi-objective genetic algorithm. In their formulation a fixed number of sensors had to be placed in order to maximize the coverage. In some applications most important is the network energy. Hernandez and Blum in [7] are proposed ACO algorithm and in [17] is proposed evolutionary algorithm for this variant of the problem. In [6] is proposed ACO algorithm taking in to account only the number of the sensors. In [10] is proposed several evolutionary algorithms to solve the problem. In [9] is proposed genetic algorithm which achieves similar solutions as the algorithms in [10], but it is tested on small test problems.

The rest of the paper is organized as follows. In Section 2 the WSN is described and the layout problem is formulated. Section 3 presents the ACO algorithm. In Section 4 the experimental results obtained are shown. Finally, several conclusions are done in Section 5.

2. WSN Layout Problem

WSN consists of spatially distributed autonomous sensors to cooperatively monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion or pollutants [1,9]. The development of wireless sensor networks was motivated by military applications such as battlefield surveillance and are now used in many industrial and civilian application areas, including industrial process monitoring and control, machine health monitoring, environment and habitat monitoring, health care applications, home automation, and traffic control [8].

A sensor node might vary in size from that of a box to the size of a grain of dust [8]. The cost of sensor nodes is similarly variable, ranging from hundreds of dollars to a few pennies, depending on the complexity required of individual sensor nodes [8]. Size and cost constraints on sensor nodes result in corresponding constraints on resources such as energy, memory, computational speed and bandwidth [8].

Each sensor node sense an area around itself called its sensing area. A parameter called sensing radius determines the sensitivity range of the sensor node and thus the sensing area. The nodes communicate among themselves using wireless communication links. These links are determined by a communication radius. A special node in the WSN called High Energy Communication Node (HECN) is responsible for external access to the network. Therefore, every sensor node in the network must have communication with the HECN. Since the communication radius is often much smaller than the network size, direct links are not possible for peripheral nodes. They transmit their data by other nodes which are closer to the HECN.

The WSN layout problem amounts to deciding the geographical position of the sensor nodes. In our formulation, a non-fixed amount of sensor nodes has to be placed in a terrain providing full sensitivity coverage with a minimal number of sensors. The positions of the nodes have to be chosen in a way that minimizes the energy of spent in communications by any single node, while keeps the connectivity of the network. Minimal number of sensors means cheapest network for constructing. Minimal energy means cheapest network for exploitation. The energy of the network defines the lifetime of the network, how frequently the batteries need to be replaced. These are opposed objectives since the more nodes there are the lesser share of retransmissions they bear and in opposite, when we try to decrease the energy consumption normally we include nodes.

In order to determine the energy spent by communications, the number of trans-missions every node performs is calculated. The WSN operates by rounds: In a round, every node collects data from its measurements and sends it to the HECN. Every node transmits the information packets to the neighbour that is closest to the HECN or the HECN itself if it is within the communication range. When several neighbours are tied for the shortest distance

from the HECN, the traffic is distributed among them. That is, if a node has n neighbours tied for shortest distance from HECN, each one receives $1/n$ of its traffic load. Therefore, every node has a traffic load equal to 1 (corresponding to its own sent data) plus the sum of all traffic loads received from neighbours that are farther from the HECN. The sensing area of the WSN is the union of the individual areas of all nodes. The number of sensor nodes must be kept as low as possible, since using many nodes represents a high cost of the network, possibly influences of the environment and also provokes a probability of detection (when stealth monitoring is designed). The objective of this problem is to minimize network energy and the number of sensors deployed while the area is fully covered and connected.

3. ACO for WSN Layout Problem

The WSN Layout problem is NP-hard. The metaheuristic methods are most appropriate for solving NP optimization problems. Many of the existing solutions of WSN Layout problem come from the field of Evolutionary Computation [2][11]. We propose ACO algorithm. The ACO is one of the most successful metaheuristic. The idea for ACO comes from real ant behaviour. Real ants foraging for food lay down quantities of pheromone marking the path that they follow. An ant encountering a previously laid pheromone will detect and decide to follow it with high probability thus taking more informed actions based on the experience of previous ants. The repetition of the above mechanism represents the auto-catalytic behaviour of real ant colony where the more the ants follow a trail, the more attractive that trail becomes. Thus ants can find the shortest path between the nest and sours of the food. The problem is represented by graph and the ants walk on the graph to construct solutions. The solution is represented by a path or by tree in the graph. After initialization of the pheromone trails, ants construct feasible solutions, starting from random nodes. Then the pheromone trails are updated. At each step ants compute a set of feasible moves and select the best one to carry out the rest of the tour. The transition probability p_{ij} , to choose the node j when the current node is i , is based on the heuristic information η_{ij} and on the pheromone trail level τ_{ij} of the move, where $i, j = 1, \dots, n$.

$$p_{ij} = (\tau_{ij} \eta_{ij}) / \sum_k \tau_{ik} \eta_{ik}, \quad (1)$$

where the sum is on the possible k .

The higher value of the pheromone and the heuristic information, the more profitable is to select this move. The pheromone corresponds to the global memory of the ants, their experience in problem solving from previous iterations. The heuristic information is apriority knowledge for the problem which is used to manage the search process to improve the solution. In the beginning, the initial pheromone level is set to a small positive constant value τ_0 and then ants update this value after completing the construction stage [3]. ACO algorithms adopt different criteria to update the pheromone level. In our implementation we use MAX-MIN Ant System (MMAS) [14], which is one of the best ant approaches. The main feature of MMAS is using a fixed upper bound τ_{\max} and a lower bound τ_{\min} of the pheromone trails. Thus the accumulation of big amounts of pheromone by part of the possible movements and repetition of same solutions is partially prevented on one side, and the amount of the pheromone to decrease a lot of and to become close to zero and the element to be unused is partially prevented in another.

The pheromone trail update rule is given by:

$$\tau_{ij} \leftarrow \rho \tau_{ij} + \Delta \tau_{ij}, \quad (2)$$

where $\Delta \tau_{ij} = 1/C(V_{\text{best}})$ if (i, j) is from the best solution and $\Delta \tau_{ij} = 0$ otherwise. The V_{best} is the iteration best solution and $i, j = 1, \dots, n$, ρ is from the interval $[0, 1]$ models evaporation in the nature. $C(V)$ is the objective function.

The WSN layout problem is NP-hard multi-objective problem. We simplify it converting to mono-objective. The new objective function is as follows:

$$C(V_k) = \lambda f_1(V_k)/\max f_{11} + (1-\lambda)f_2(V_k)/\max f_{21} \quad (3)$$

where V_k is the solution constructed by the ant k and $f_1(V_k)$ and $f_2(V_k)$ are the number of sensors and energy corresponding to the solution V_k . f_{11} is the maximal value of f_1 from the first iteration, respectively f_{21} . Dividing f_1 by $\max f_{11}$ and f_2 by $\max f_{21}$ is a sort to normalize the values of f_1 and f_2 respectively, because they are rather different. The parameter λ shows the influence of the objective functions f_1 and f_2 in the new objective function. To avoid stagnation of the search, the range of possible pheromone values on each movement is limited to an interval $[\tau_{\min}, \tau_{\max}]$. τ_{\max} is an asymptotic maximum of τ_{ij} and $\tau_{\max} = 1/(1-\rho)C(V^*)$, while $\tau_{\min} = 0.087 \tau_{\max}$. Where V^* is the optimal solution, but it is unknown, therefore we use V_{best} , the current best value of the objective function. One of the crucial point of the ACO algorithms is construction of the graph of the problem. In our implementation the WSN layout problem is represented by two graphs, it is one of our contributions. The terrain is modeled by grid $G = \{g_{ij}\}_{N \times M}$, where M and N are the size of the sensing region. By the graph G we calculate the coverage of the terrain. We use another graph $G1_{N1 \times M1}$, on which nodes we map the sensors, where $N1 \leq N$ and $M1 \leq M$. The parameters $M1$ and $N1$ depend of the sensing and communication radius. Thus we decrease the number of calculations the algorithm performs, respectively the running time. The pheromone is related with location sites $Ph = \{ph_{ij}\}_{N1 \times M1}$, the initial pheromone can be a small value, for example $1/n_{\text{ants}}$. The point, where the HECN is located, is included in the solutions like first point. Every ant starts to create the rest of the solution from a random node which communicates with first one. The ant chooses the next position by the ACO probabilistic rule. It chooses the point having the highest probability. If there is more than one point with same probability, the ant chooses one of them randomly.

The construction of heuristic information is another crucial point of the ACO algorithm. The heuristic information needs to manage the ants to look for better solutions and to avoid unfeasible solutions. The second contribution of this paper is proposition of appropriate heuristic information. Our heuristic information is a product of three parameters as follows:

$$\eta_{ij} = s_{ij} l_{ij} (1 - b_{ij}) \quad (4)$$

s_{ij} is the number of points which the new sensor will cover, $l_{ij}=1$ if exist communications with other nodes and $l_{ij}=0$ otherwise. b_{ij} is the solution matrix and the matrix element $b_{ij}=1$ when there is sensor on the node (i,j) of the graph $G1$, otherwise $b_{ij}=0$. With s_{ij} we try to locally increase the covered points, more new covered point's leads eventually to less number of sensors. With l_{ij} we guarantee that all sensors will be connected; with rule $(1-b_{ij})$ we guarantee that the position is not chosen yet and no more than one sensor will be mapped on the same node of the graph $G1$. When $p_{ij}=0$ for all values of i and j the search stops. Thus, the construction of the solution stops if no more free positions, or all points are covered or new communication is impossible.

4. Experimental Results

We will contribute with this work to improve the state-of-the-art of the use of metaheuristics for solving the WSN layout problem. With our algorithm we can solve WSN layout problem on any rectangular area and the HECN can be fixed on any point on the area. We test our algorithm on WSN problem instance where a terrain of 500x 500 points has to be covered using sensors with coverage and communication radii covered 30 points in graph G . In ideal conditions only 89 sensors would be necessary, but it is impossible due to their geometrical shape (circle). Therefore, the expected minimum number of nodes for full-coverage is higher than 89. Thus the graph G consists of 500x500 nodes. When we apply our ACO algorithm on this example the graph $G1$ consists of 50x50 nodes which are 100 times less than the graph G . The number of nodes of the graph $G1$ is proportional to the number of the nodes of the graph G . Thus the nodes of the graph $G1$ are mapped on nodes of the graph G and coverage and communication radii cover 3 points from the graph $G1$. In our example the HECN is fixed in the centre of the area.

An example of solution that achieves full coverage of the region is a square grid formed by the sensors separated by 30 points of graph G. Starting at the HECN, 250 points have to be covered to each side of the terrain, requiring 8 sensors. Therefore the grid has 17 (8+8+1) rows and 17 columns, thus 289 sensors, including the HECN. In this symmetrical configuration there are four nodes directly connected to the HECN, so the complete traffic of the network 288 messages per round is evenly divided among them. This results in the most loaded nodes having a load of 72 messages. So this candidate solution obtains (288, 72).

After several runs of the algorithm we specify the most appropriate values of its parameters. We apply MAX-MIN ant algorithm with the following parameters: $\rho=0.5$, the number of used ants is 4 and the maximum number of iterations is 5. The parameter λ has values from the set {0.125, 0.250, 0.375, 0.5, 0.625, 0.750, 0.875}. In Table 1 are reported non-dominated solutions for every value of λ . In Table 2 are reported best found results (with respect to the sensors and with respect to the energy) achieved by several metaheuristic methods [10]. We compare our ACO algorithm results with results obtained by the evolutionary algorithms in [10] and the symmetric solution. These evolutionary algorithms perform like multi-objective and reports non-dominated solutions. A solution dominates other if all components of the first are not worse than the second. A solution is non-dominated if and only if no other solution dominates it.

Table 1. Non-dominated ACO solutions

λ	Non dominated solutions
0.125	(238,49) (235,51) (229,58) (227,60)
0.250	(238,49) (235,51) (228,58)
0.375	(238,49) (235,51) (229,57) (227,58) (226,60)
0.500	(238,49) (233,56) (229,57) (228,58) (226,60)
0.625	(238,49) (235,51) (229,57) (227,58) (226,60)
0.750	(238,49) (235,51) (229,57) (227,58) (226,60)
0.875	(238,49) (230,51) (225,57)

Table 2. WSN solutions

Algorithm	Min sensors	Min energy
Symmetric	(288,72)	(288,72)
MOEA	(260,123)	(291,36)
NSGA-II	(262,83)	(277,41)
IBEA	(265,83)	(277,41)

We perform 30 independent runs of the ACO algorithm and the achieved numbers of sensors are in the interval [225, 247]. Thus the worst number of sensors achieved by ACO algorithm is less than the best number of sensors achieved by other mentioned algorithms. All non-dominated ACO solutions dominate the symmetric solution. Let compare achieved solutions with minimal number of sensors. The solutions achieved by mentioned evolutionary algorithms have very high energy more than symmetric solution. Thus they are not good solutions. The ACO solution with minimal number of sensors dominates other solutions with minimal number of sensors and symmetric solution. Thus it is a good solution. Let compare solutions with minimal energy achieved by mentioned algorithms. MOEA algorithm achieves solution with very small value for energy, but too many sensors, more than symmetric solution, thus it is not good solution. Other two evolutionary algorithms achieve solutions with a less energy than symmetric and a little bit less number of sensors. Thus they are not bed solutions. The ACO solution dominates symmetric one. Its energy is a little bit more than the evolutionary algorithms, but the number of sensors is much less. Comparing the influence of the parameter λ , the best value for λ is $\lambda = 0.875$, because in this case the algorithm achieves solutions, which dominates the solutions, achieved by other values of the λ . We can conclude that our ACO algorithm achieves very encouraging solutions.

5. Conclusion

We have defined wireless sensor networks layout problem with its connectivity constraint. We simplify the problem converting it from multi-objective to mono-objective. The new objective function is a combination of the number of sensors and energy of the network with weights. We learn the influence of the two objective functions on the result quality. The best result we achieved when the energy is 7 times less weighted than the number of sensors. We decrease the number of computations describing the problem by two graphs (square grids) with different size. The results of the experiments indicate very encouraging performance of ACO algorithm.

Acknowledgment

This work has been partially supported by the Bulgarian National Scientific Fund under the Grants DID 02/29 and DTK 02/44.

References

1. Akulidiz I.F., Su W., Sankarasubramaniam Y., Cayrci E., 2001, Wireless Sensor Networks: A Survey, Computer Networks, Vol. 38(4), Elsevier, pp. 393-422.
2. Alba E., Molina G., 2008, Optimal Wireless Sensor Layout with Meta-heuristics: Solving a Large Scale Instance, Large-Scale Scientific Computing, LNCS, Vol. 4818, Springer, pp. 527-535.
3. Bonabeau E., Dorigo M., Theraulaz G., 1999, Swarm Intelligence: From Natural to Artificial Systems, Oxford University Press.
4. Cahon S., Melab N., Talbi El-G., Paradiseo: A Framework for the Reusable Design of Parallel and Distributed Metaheuristics, J. of Heuristics, Vol. 10(3), pp. 357-380.
5. Deb K., Pratap A., Agrawal, Meyarivan T., 2002, A Fast and Elitist Multiobjective Genetic Algorithm: Nsga-ii, IEEE Transactions on Evolutionary Computation, Vol. 6(2), pp. 182-197.
6. Fidanova S., Marinov P., Alba E., 2010, ACO for Optimal Sensor Layout, In Proc. of Int. Conf. on Evolutionary Computing, Valencia, Spain, Joaquim Filipe and Janus Kacprzyk eds., SciTePress-Science and Technology Publications Portugal, ISBN 978-989-8425-31-7, pp. 5-9.
7. Hernandez H., C. Blum, 2011, Minimum Energy Broadcasting in Wireless Sensor Networks: An ant Colony Optimization Approach for a Realistic Antenna Model, J. of Applied Soft Computing, Vol. 11(8), pp. 5684-5694.
8. Jourdan D.B., 2000, Wireless Sensor Network Planning with Application to UWB Localization in GPS-denied Environments, Massachusetts Institute of Technology, PhD thesis.
9. Konstantinidis A., Yang K., Zhang Q., Zainalipour-Yazti D., 2010, A multi-objective Evolutionary Algorithm for the deployment and Power Assignment Problem in Wireless sensor Networks, J. of Computer networks, Vol. 54(6), pp. 960-976.
10. Nemeroff J., Garcia L., Hampel D., DiPierro S., 2001, Application of Sensor Network Communications, IEEE Military Comm. Conf., pp. 336-341.
11. Paek J., Kothari N., Chintalapudi K., Rangwala S., Govindan R., 2005, The Performance of a Wireless Sensor Network for Structural Health Monitoring, In Proc. of 2nd European Workshop on Wireless Sensor Networks, Istanbul, Turkey,.
12. Pottie G.J., Kaiser W.J., 2000, Embedding the Internet: Wireless Integrated Network Sensors, Communications of the ACM, Vol. 43(5), pp. 51-58.
13. Romer K., Mattern F., 2004, The Design Space of Wireless Sensor Networks, IEEE Wireless Communications, Vol. 11(6), ISSN 1536-1284, pp. 54-61.
14. Stutzle T., Hoos H.H., 2000, MAX-MIN Ant System, Future Generation Computer Systems, Vol. 16, 2000, pp. 889-914.
15. Werner-Allen G., Lorinez K., Welsh M., Marcillo O., Jonson J., Ruiz M., Lees J., 2006, Deploying a Wireless Sensor Network on an Active Volcano, IEEE Internet Computing, Vol. 10(2), pp. 18-25.

16. Wolf S., Mezz P., 2008, Evolutionary Local Search for the Minimum Energy Broadcast Problem, in C. Cotta, J. van Hemezel (Eds.), VOCOP 2008, Lecture Notes in Computer Sciences, Vol. 4972, Springer, Germany, pp. 61-72.
17. Xu Y., Heidemann J., Estrin D., 2001, Geography Informed Energy Conservation for Ad Hoc Routing, Proceedings of the 7th ACM/IEEE Annual International Conference on Mobile Computing and Networking, Italy, pp. 70-84.
18. Yuce M.R., Ng S.W., Myo N.L., Khan J.Y., Liu W., 2007, Wireless Body Sensor Network Using Medical Implant Band, Medical Systems, Vol. 31(6), pp. 467-474.
19. Zitzler E., Knzli S., 2004, Indicator-based Selection in Multiobjective Search, PPSN'04, LNCS, Vol. 3242, Springer, pp. 832-838.

Application of Simulated Annealing and Threshold Accepting for a Parameter Tuning Problem

Olympia Roeva¹, Tsonyo Slavov²

¹Institute of Biophysics and Biomedical Engineering - BAS, Sofia, Bulgaria,

²Technical University of Sofia, Sofia, Bulgaria

¹olympia@biomed.bas.bg, ²ts_slavov@tu-sofia.bg

Abstract: In this paper two local search trajectory methods – Simulated Annealing (SA) and Threshold Accepting (TA) were applied for a PID controller parameter tuning. The controllers are used to control feed rate and to maintain glucose concentration at the desired set point for an *E. coli* fed-batch cultivation process. A system of nonlinear differential equations was used to model the biomass growth and substrate utilization. The SA and TA adjustments were done based on several pre-tests according to the optimization problem considered here. Simulation results indicate that the applied trajectory methods are effective and efficient. Good closed-loop system performance is achieved on the basis of the considered PID controllers tuning procedures.

Keywords: Simulated Annealing, Threshold Accepting, Cultivation Process, PID Controller, Parameter Tuning.

1. Introduction

Classical biotechnology is the science of production of human-useful processes and products under controlled conditions, applying biological agents – micro-organisms, plant or animal cells, their exo- and endo- products, e.g. enzymes, etc. [14].

Cultivation of recombinant micro-organisms e.g. *E. coli*, in many cases is the only economical way to produce pharmaceutical biochemicals such as interleukins, insulin, interferons, enzymes and growth factors. To maximize the volumetric productivities of bacterial cultures it is important to grow *E. coli* to high cell concentration. Among the different modes of operation, (batch, fed-batch and continuous), fed-batch operation is the one most often used in industry. Since both nutrients overfeeding and underfeeding is detrimental to cell growth and product formation, development of a suitable feeding strategy control is critical in fed-batch cultivation. The control strategy for substrate feed rate can be summarized in three groups: open (feedforward), closed-loop (feedback) control and mixed (feedforward-feedback). In feedback control of industrial cultivation processes, the proportional-integral-derivative (PID) controller is widely used.

Usually the PID controller is poorly tuned due to highly changing dynamics of most bioprocesses caused by the nonlinear growth of the cells and the changes in the overall metabolism. The tuning procedure is a significant challenge for the conventional optimization methods. As an alternative, metaheuristics could be applied [7][8][10].

During the last decade, metaheuristic techniques have been successfully applied to a variety of areas. Optimization heuristics are generally divided into two broad classes, namely, constructive methods and local search methods. For a long time local search has not been considered as a mature technique and only recently the method become an object of increasing interest [6]. Local search methods are divided into trajectory methods and population based methods. In this paper, we investigate some methods from the first class, namely Simulated Annealing (SA) [3][9][11] and Threshold Accepting (TA) [5][15].

In the literature, there are results showing different strategies based on SA for the optimal tuning of a PID controller considering linear systems [2][4][10][16]. The researches so far demonstrate that SA provides an optimal tuning of the PID controller. However, no results

have been found about using SA or TA for nonlinear control systems design, such as the considered here problem.

The authors have already successfully applied Genetic Algorithms (GA) to the optimal tuning of a PID controller concerning nonlinear system [13]. The main contribution of the present paper is the application of the trajectory methods such as SA and TA, specialized to solve PID controller parameter tuning for glucose control of a nonlinear fed-batch cultivation process for the first time.

2. Trajectory Methods

The definition of neighbourhood is central to trajectory methods and it generally depends on the problem under consideration. Finding an efficient function that leads to high quality local optima can be challenging [6].

2.1. Simulated Annealing Algorithm

SA is a local search method where the search mechanism is modelled on the Metropolis et al. [11] algorithm and the principles of thermodynamic annealing. Kirkpatrick et al. [9] and Černý [3] were the first to follow such a technique to solve combinatorial optimization problems.

SA can deal with arbitrary systems and cost functions. It statistically guarantees finding an optimal solution; guarantees a convergence upon running a sufficiently large number of iterations and is relatively easy to code, even for complex problems. This makes SA an attractive option for optimization problems.

The structure of the SA algorithm, applied here to PID controller tuning for control of the considered nonlinear process, is demonstrate by the pseudo code below:

Simulated Annealing Algorithm

1. Generate initial solution x^c ,
initialize R_{\max} and T ;
2. **for** r to R_{\max} **do**
3. **while** stopping criteria not met **do**
4. Compute $x^n \in \mathbb{N}(x^c)$
 (neighbour to current solution)
5. Compute $\Delta = f(x^n) - f(x^c)$ and
 generate u (uniform random variable)
6. **if** $(\Delta < 0)$ or $e^{-\Delta/T} > u$ **then** $x^c = x^n$
7. **end while**
8. Reduce T
9. **end for**

2.2. Threshold Accepting Algorithm

The TA method, suggested by Dueck and Scheuer [5] is a less mentioned method. TA is a deterministic analogue of SA. The main difference between these two methods is the level of acceptance of a lower quality solution et each step. In a TA method, this type of decision is taken in a deterministic manner, without having to use the principles of thermodynamic annealing.

The structure of the applied here TA algorithm is demonstrated by the pseudo code below:

Threshold Accepting Algorithm

1. Generate initial solution x^c ,
initialize τ ;
2. **for** r to 0 **do**
3. **while** stopping criteria not met **do**
4. Compute $x^n \in \aleph(x^c)$
5. Compute $\Delta = f(x^n) - f(x^c)$
6. **if** $\Delta < \tau$ **then** $x^c = x^n$
7. **end while**
8. Reduce τ
9. **end for**

In this algorithm, the sequence of temperatures T is replaced by a sequence of thresholds τ .

SA, as well as TA, are generic techniques that may require excessive computing times to achieve good solutions and are highly problem dependent, as far as tuning of algorithms parameters is concerned. Parameters of the SA and TA were tuned, based on several pre-tests according to the problem considered here. After tuning procedures, the main algorithm parameters are set to the optimal settings. The reannealing interval is set to 100. The initial temperature or threshold is set to 100. The maximum number of objective function evaluations is set to 9000. The number of iterations, over which average change in fitness function value at current point is less than termination tolerance on function value, is set to 1500. The termination tolerance on function value is 10^{-6} .

3. Problem Formulation

In the considered here glucose concentration control systems of an *E. coli* fed-batch cultivation process a universal digital PID controller is used. The structure of the PID control system is shown on Figure 1.

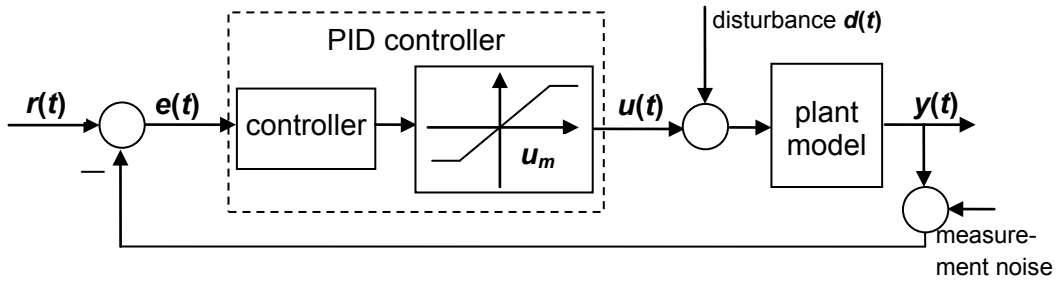


Figure 1: Structure of the control system

The error signal $e(t)$ is used to generate the P, I, and D modes, with the resulting signals weighed and summed to form the control signal $u(t)$ applied to the plant model (1) – (5) [12]:

$$\begin{cases} \dot{\mathbf{x}}(t) = f(\mathbf{x}, F) + \boldsymbol{\eta}(t) \\ S(t) = \mathbf{H}\mathbf{x}(t) + \xi(t) \end{cases} \quad (1)$$

$$\mathbf{x}(t) = [X(t) \quad S(t) \quad V(t) \quad \mu_{\max}(t)]^T \quad (2)$$

$$f(\mathbf{x}, F) = \begin{bmatrix} \mu(t) X(t) - \frac{F(t)}{V(t)} X(t) \\ -\frac{1}{Y_{S/X}} \mu(t) X(t) + \frac{F(t)}{V(t)} (S_{in} - S(t)) \\ F(t) \\ 0 \end{bmatrix}, \quad (3)$$

$$\mu(t) = \mu_{\max}(t) \frac{S(t)}{k_S + S(t)} \quad (4)$$

$$\mathbf{H} = [0 \quad 1 \quad 0 \quad 0], \quad (5)$$

$$\boldsymbol{\eta}(t) = [\eta_{\gamma_X}(t) \quad \eta_{\gamma_S}(t) \quad 0 \quad \eta_{\mu_{\max}}(t)]^T, \quad (6)$$

where X is the biomass concentration, $\text{g} \cdot \text{l}^{-1}$; S – substrate (glucose) concentration, $\text{g} \cdot \text{l}^{-1}$; F – feed rate, $\text{l} \cdot \text{h}^{-1}$; V – bioreactor volume, l ; S_{in} – substrate concentration of the feeding solution, $\text{g} \cdot \text{l}^{-1}$; μ_{\max} – maximum growth rate, h^{-1} ; k_S – saturation constant, $\text{g} \cdot \text{l}^{-1}$; $Y_{S/X}$ – yield coefficient, –; η_{γ_X} – biomass concentration process noises, $\text{g} \cdot \text{l}^{-1}$; η_{γ_S} – substrate concentration process noises, $\text{g} \cdot \text{l}^{-1}$; $\eta_{\mu_{\max}}$ – maximum growth rate process noises, h^{-1} ; $\xi(t)$ – measurement noise, $\text{g} \cdot \text{l}^{-1}$.

The model inaccuracy is modelled via zero mean white Gaussian noises. The corresponding variances are [1]:

$$D_{\eta_{\gamma_X}} = 0.001 \text{ g}^2 \cdot \text{l}^{-2} \cdot \text{h}^{-1}, \quad D_{\eta_{\gamma_S}} = 0.001 \text{ g}^2 \cdot \text{l}^{-2} \cdot \text{h}^{-1}, \quad D_{\eta_Q} = 0 \text{ l}^2 \cdot \text{h}^{-3} \text{ and } D_{\eta_{\mu_{\max}}} = 0.05 \text{ l} \cdot \text{h}^{-3}.$$

The initial process conditions are [1]:

$$t_0 = 6.68 \text{ h}, \quad S_{in} = 100 \text{ g} \cdot \text{l}^{-1}, \quad X(t_0) = 1.25 \text{ g} \cdot \text{l}^{-1} \text{ and } S(t_0) = 0.8 \text{ g} \cdot \text{l}^{-1}.$$

The mathematical description of discrete-time universal PID controller is:

$$u(k) = u_p(k) + u_i(k) + u_d(k), \quad (7)$$

$$u_p(k) = K_p [br(k) - y(k)], \quad (8)$$

$$u_i(k) = u_i(k-1) + b_{i1} [r(k) - y(k)] + b_{i2} [r(k-1) - y(k-1)], \quad (9)$$

$$u_d(k) = a_d u_d(k-1) + b_d [cr(k) - cr(k-1) - y(k) + y(k-1)], \quad (10)$$

where $r(k)$ is a reference signal, $y(k)$ – output signal, $u(k)$ – control signal, K_p – proportional gain, T_i – integral time, T_d – derivative time, T_d/N – time constant of first-order low pass filter, T_0 – sample time and $b_{i1} = K_p T_0 / T_i$, $b_{i2} = 0$, $a_d = T_d / (T_d + NT_0)$, $b_d = K_p T_d N / (T_d + NT_0)$.

To evaluate the significance of the tuning procedure and controller performance an integrated square error (I_{ISE}) is used as an objective function:

$$I_{ISE} = \int_0^{t_{end}} e(t)^2 dt \quad (11)$$

where the error e is the difference between the set point and the estimated substrate concentration ($S_{SP} - S$), t – time, t_{end} – end time of the cultivation.

For the considered here *E. coli* cultivation process, the desired set point is at $S_{SP} = 0.1 \text{ g} \cdot \text{l}^{-1}$ glucose concentration [1].

4. Numerical Results and Discussion

A series of tuning procedures for PID controllers tuning, using SA and TS, are performed. Computer specifications to run all optimization procedures are Intel® Core™i5-2320 CPU @ 3.00GHz, 8 GB Memory(RAM), Windows 7 (64bit) operating system and Matlab 7.5.0 environment. Because of the stochastic characteristics of the applied algorithms, a series of 30 runs for each algorithm were performed. For comparison of the SA and TA, the mean and the best results of the 30 runs, for the number of function evaluations (N_{FE}), total time for the solver to run and objective function value I_{ISE} are observed.

The range of the tuning parameters is considered as follows:

$$K_p \in [0, 2], T_i \in [0, 1], T_d \in [0, 0.1].$$

After several runs the range for the parameters is specified to:

$$K_p \in [0.4, 2], T_i \in [0.005, 1] \text{ and } T_d \in [0.003, 0.1].$$

The coefficients b , c and N are set to the following values [13]: $b = 1$, $c = 1$ and $N = 30$.

As a result of the SA and TA tuning, the optimal PID controllers settings are obtained. The numerical values of the controllers parameters, objective function values, total times and number of function evaluations are presented in Table 1 and Table 2, respectively for SA and TA algorithm.

Table 1: Results from PID controller tuning using SA

Parameter	Value	
	Mean	Best
K_p	0.4002	0.4000
T_i	0.9933	0.9990
T_d	0.0030	0.0030
I_{ISE}	110.4504	110.4503
total time, s	93.4555	57.2524
N_{FE}	2576	1597

Table 2: Results from PID controller tuning using TA

Parameter	Value	
	Mean	Best
K_p	0.4007	0.4000
T_i	0.9536	0.9994
T_d	0.0030	0.0030
I_{ISE}	110.4504	110.4503
total time, s	94.4321	60.9184
N_{FE}	2236	1542

SA and TA algorithms are found to be quite efficient from the computational experiments. Results from SA and TA PID controller tuning show that the algorithms produce the same estimations with a very high percentage coincidence. The numerical values of the controllers parameters are similar for each algorithm. The obtained values of I_{ISE} for both algorithms are representative and sophisticated controller performance indices. Both algorithms achieved the same objective function value – 110.4503. No essential differences are observed about the total time and number of function evaluations of the two algorithms. For all the test instances both algorithms obtain optimal solutions for a short period of computational time.

Analysing the obtained results, it can be concluded that SA and TA algorithms are functionally and statistically equal.

In comparison to GA performance, it is observed that SA and TA are superior to GA with respect to computational time and number of function evaluations. Applying GA, based on [13], the following mean results are obtained:

- objective function value – 110.4510;
- total time – 187.87 s.;
- N_{FE} – 5200.

In the next figures, some results of control system performance for *E. coli* fed-batch cultivation process are presented. In Figure 2 the control variable (glucose concentration) is presented. Due to the obtained similar results, only TA tuned PID controller performance is displayed. It is seen that for a short time the controller sets the control variable and keeps it at the desired set point of $0.1 \text{ g}\cdot\text{l}^{-1}$ during the process.

In Figure 3, feed rate profile during the *E. coli* fed-batch cultivation process is presented. Optimal controller parameters lead to a high biomass concentration in the end of the cultivation process – $17.2 \text{ g}\cdot\text{l}^{-1}$ (see Figure 4).

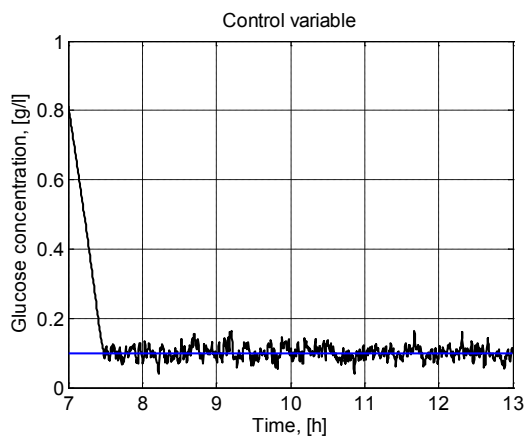


Figure 2: Glucose concentration during the process

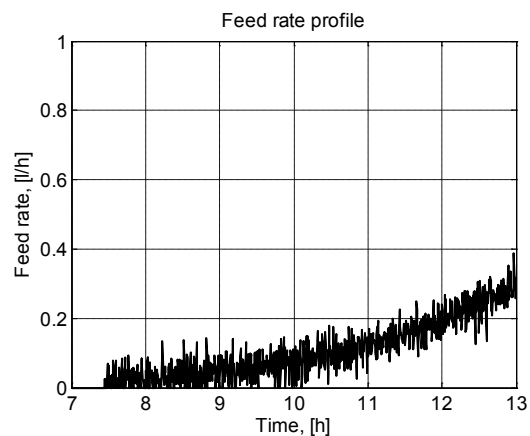


Figure 3: Feed rate profile during the process

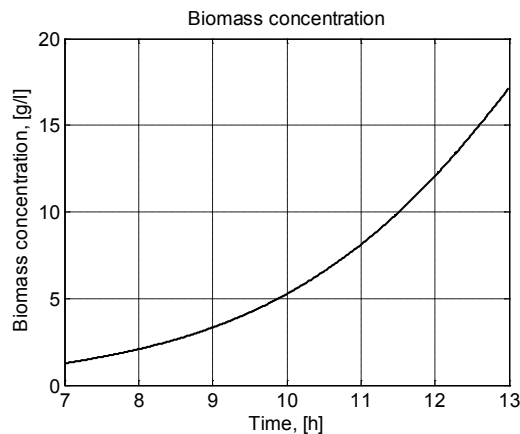


Figure 4: Biomass concentration during the process

5. Conclusion

This paper presents an optimal tuning of PID controller using two trajectory methods, namely, SA and TA. The controller is used to control feed rate and to maintain glucose concentration at the desired set point for an *E. coli* fed-batch cultivation process. The mathematical model of the cultivation process is represented by the dynamic mass balance equations for main process variables – biomass and substrate concentration.

A series of tuning procedures for PID controllers tuning, using both techniques, are performed. The SA and TA algorithms' parameters are problem-oriented and specifically chosen to achieve an adequate and accurate decision. As a result, a set of optimal SA and TS PID controllers parameters (K_p , T_i and T_d) are obtained. For a short time, the controllers set the control variable and maintain it at the desired set point during the cultivation process. Thus, a good closed-loop system performance is achieved.

Based on the observed results, it could be concluded that both algorithms, SA and TA, show identical performance for PID controller parameter tuning of the considered nonlinear control system.

In summary, it is shown that the PID controller tuning using SA and TA can be considered as an effective methodology for achievement of high quality and better performance of the designed control system for cultivation processes.

Acknowledgments

This work is partially supported by the National Science Fund Grants DMU 02/4 "High Quality Control of Biotechnological Processes with Application of Modified Conventional and Metaheuristics Methods" and DID 02-29 "Modeling Processes with Fixed Development Rules (ModProFix)".

References

1. Arndt M., Hitzmann B., 2001, Feed Forward/Feedback Control of Glucose Concentration during Cultivation of *Escherichia coli*, In: 8th IFAC Int. Conf. on Comp. Appl. in Biotechn., Canada, pp. 425-429.
2. Asharf S., Kumar S., 2008, PUMA 560 Optimal Trajectory Control using Genetic Algorithm, Simulated Annealing and Generalized Pattern Search Techniques. International Journal of Electrical, Computer, and Systems Engineering, Vol. 1(2), pp. 71-80.
3. Černý V., 1985, Thermodynamical Approach to the Travelling Salesman Problem: An Efficient Simulation Algorithm, Journal of Optimization Theory and Applications, Vol. 45, pp. 41-51.
4. Dashti M. J., Kambiz S. G., Seyedkashi S. M. H., Behnam T. M., 2010, Novel Simulated Annealing Algorithm in Order to Optimal Adjustment of Digital PID Controller, In: 11th Int. Conf. on Control Automation Robotics & Vision (ICARCV), 7-10 Dec. 2010, pp. 1766-1771.
5. Dueck G., Scheuer T., 1990, Threshold accepting: A general purpose algorithm appearing superior to simulated annealing, Journal of Computational Physics, Vol. 90, pp. 161-175.
6. Gilli M., Winker P., 2009, Heuristic Optimization Methods in Econometrics, In Handbook of Computational Econometrics edited by D. Belsley, E. Kontoghiorghe, John Wiley & Sons, Ltd.
7. Gundogdu O., 2005, Optimal-tuning of PID Controller Gains using Genetic Algorithms, Journal of Engineering Sciences, Vol. 11(1), pp. 131-135.
8. Kim J. S., Kim J.-H., Park J.-M., Park S.-M., Choe W.-Y., Heo H., 2008, Auto Tuning PID Controller based on Improved Genetic Algorithm for Reverse Osmosis Plant. World Academy of Science, Engineering and Technology, Vol. 47, pp. 384-389.
9. Kirkpatrick S., Gelatt C. D. Jr., Vecchi M. P., 1983, Optimization by Simulated Annealing, Science, Vol. 220, pp. 671-680.

10. Kumar S. M. G., Rakesh B., Anantharaman N., 2010, Design of Controller using Simulated Annealing for a Real Time Process, International Journal of Computer Applications, Vol. 2, pp. 1053-1368.
11. Metropolis N., Rosenbluth A., Rosenbluth M., Teller A., Teller M., 1953, Equation of State Calculations by Fast Computing Machines, Journal of Chemical Physics, Vol. 21, pp. 1087-1092.
12. Roeva O., 2008, Parameter Estimation of a Monod-type Model based on Genetic Algorithms and Sensitivity Analysis, Lecture Notes on Computer Science, Springer-Verlag Berlin Heidelberg, Vol. 4818, pp. 601-608.
13. Slavov Ts., Roeva O., 2011, Genetic Algorithm Tuning of PID Controller in Smith Predictor for Glucose Concentration Control, International Journal Bioautomation, Vol. 15(2), pp. 101-114.
14. Viesturs U., Karklina D, Ciprovica I., 2004, Bioprocess and Bioengineering, Jeglava.
15. Winker P., 2001, Optimization Heuristics in Econometrics: Applications of Threshold Accepting, Wiley, Chichester.
16. Yacoub A. H. Y., Buyamin S., Wahab N. A., 2011, Integral Time Absolute Error Minimization for PI Controller on Coupled-tank Liquid Level Control System based on Stochastic Search Techniques, Journal Technology, Vol. 54, pp. 381-402.

Special Session

Wireless Networks - Collaborative Information and Control Systems

Keynote: Collaboration in Wireless Networks – Towards New Paradigm in Building Intelligent Traffic Systems

Dr Evtim Peytchev, Nottingham Trent University

Collaboration in Wireless Networks - Towards New Paradigm in Building Intelligent Traffic Systems

Evtim Peytchev,
Nottingham Trent University - School of Science and Technology
Computing and Informatics Building, Clifton Lane, Nottingham, NG11 8NS, UK
evitm.peytchev@ntu.ac.uk

Abstract

One of the most important areas of investigation nowadays is the area of Intelligent Traffic Systems – ITS. The importance of such systems can have many different aspects, but the most important ones arguably are the aspect of keeping pollution to a minimum while reducing journey times and saving time and money of the commuters and, at the same time, the aspect of utilising to a maximum the existing road capacity and taking into account all commuters preferences.

In this plenary lecture I will be revisiting the topic of urban traffic control, look into the past, present and the future of the developments in this area. The lecture will present an informative view for a possible (and likely) future development of the traffic control in the cities based on new paradigm – all cars take part in the traffic information collection and integration and subsequently in the traffic control through cooperative new generation algorithms implemented and deployed within every cars' control device.

Historically the traffic control gradually moved from fixed timings for the traffic lights (inefficient and not taking into account current traffic flows) to variable demand-driven timings – the SCOOT system. The SCOOT system is the first attempt at automating in real-time the control of urban traffic – design circa 1980. SCOOT stands for Split, Cycle and Offset Optimisation Technique and it is based on TRANSYT – a macro simulation traffic model developed at the Transportation Research Laboratory (TRL) in London, UK. It has two main deficiencies:

- Limited reach – too expensive to deploy for the whole city
- Limited efficiency – the commuters do not participate in the process

As a result these two deficiencies have been addressed over the years through new approaches to help with traffic detection and with traffic control generation and implementation.

Addressing traffic detection new techniques include:

- Inductive loop detection
- Video vehicle detection
- Bluetooth detection
- Floating car data/floating cellular data
- Public transport input

Addressing the traffic control efficiency aspect is its infancy. We are yet to understand how different parts of the traffic (including commuters) interact with each other and we are yet to include all commuters in the process. Newly formulated approaches in this respect include:

- Cooperation for traffic conditions detection through wireless ad-hoc networks
- Cooperation for traffic presence detection through wireless ad-hoc networks
- Cooperation for traffic control generation through wireless ad-hoc networks – “floating” traffic control
- Cooperation for effective implementation of traffic control through (wireless) “social” networking

Conclusion:

The future of the traffic control is based on Cooperation, Cooperation and Cooperation between all participants in the traffic and based on ad-hoc wireless vehicle-to-vehicle and vehicle-to-infrastructure networking including all vehicles.

Sensing Traffic Conditions in VANET for Building New Generation in Wireless Intelligent Transportation Information Systems

Emadeddin Gamati¹, Evitm Peytchev²

Nottingham Trent University - School of Science and Technology

Computing and Informatics Building,

Clifton Lane, Nottingham, NG11 8NS, UK

¹emadeddin.gamati@ntu.ac.uk, ²evitm.peytchev@ntu.ac.uk

Abstract: Vehicle to vehicle communication (V2VC) is one of the modern approaches for exchanging and generating traffic information with (yet to be realised) potential to improve road safety, driving comfort and traffic control. In this paper, we present a novel algorithm, which is based on V2V communication, uses in-vehicle sensor information and in collaboration with the other vehicles' sensor information can detect road conditions and determine the geographical area where this road condition exists. The Traffic Condition Detection Algorithm (TCDA) – which we propose here – is simple, flexible and fast and does not rely on any kind of roadside infrastructure equipment. It will offer live road conditions information channels at – almost – no cost to the drivers and public/private traffic agencies and has the potential to become indispensable part of any future intelligent traffic system (ITS). The benefits from applying this algorithm in traffic networks are identified and quantified through building a simulation model for the widely used Network Simulator II (NS2).

Keywords: Ad Hoc Network, Vehicular Ad-Hoc Networks (VANET), Vehicular Networks, Collaborative Knowledge Generation.

1. Introduction

One of the new types of ad hoc networks, the recently emerging VANETs, are networks, which consider every vehicle as a node in the wireless network. Broadcasting in ad hoc network is an elementary operation to support many applications in VANET [1] thus encouraging vehicle manufacturers and researchers to invest more in the techniques of information gathering and dissemination in VANETs. Nowadays, so many vehicles are running on roads that serious traffic problems including traffic accidents and traffic jams are becoming a serious problem. In the US, about six millions of injuries occurred every year [5]. In 2007, more than 325k traffic accidents occurred in China, which caused more than 80k deaths and more than 380k injuries. Investigations show that most traffic accidents are collisions, however, 60% of crashes would be avoided by 0.5 sec earlier warning [9]. One obvious conclusion is that an attempt should be made to distribute the safety messages earlier to the vehicles with possibility of accidents. On the other hand, the traffic conditions of big cities become terrible, which results in time wasting, gasoline exhausting and serious air pollution. Some works should be done to make drivers know which ways can be selected to avoid traffic jams. Broadcasting in VANETs can disseminate assisting traffic condition messages to all vehicles within a certain geographical area to make drivers of vehicle control models pre-act to avoid accidents and pre-select ways to avoid traffic jams. In this mode, VANETs rely heavily on broadcast to transmit emergency message efficiently in modern road traffic environment [12].

The VANETs are considered as a specific case of MANETs, therefore they have MANETs characteristics, they are multi-hopped, decentralized and self-organized etc. In VANETs, vehicles running on the road can constitute decentralized, burst and temporary networks. The shape of the network can be determined using the different approaches. For example, the vehicle infrastructure Integration (VII) initiative in US proposes that the information about an accident should be communicated through VANET within half a second to all equipped vehicles in a 500m range [8]. Because of the shared wireless medium, a simple broadcast

scheme, which is called flooding, may lead to frequent contention and collisions in transmission among neighbours [7]. The inherent problem of the plain flooding technique is the huge amount of superfluous retransmissions that consume the limited network resources [2]. These may lead to lower reliability and more latency [11]. At an extreme condition, the channels may be blocked and broadcasts may fail, this phenomenon is called broadcast storm. In reality, traffic jams often occur in big cities at the traffic peaks, more than 1 Km long saturated traffic jams are common. The investigation shows that the traffic become saturated when the density reach 133 vehicles/Km/lane [9]. The flooding scheme is infeasible in dense VANETs, because it brings us serious contentions, redundancies and collisions.

One of the main advantages of the ad-hoc networks is the opportunity to use collaborative effort in connecting and delivering network messages as necessary [6]. This opportunity is under-utilised so far in the area of traffic control and traffic information systems where every car can be considered to be a node in an ad-hoc network [10]. Our aim is to investigate the possibility of bringing ad-hoc collaborative information generation and control into such systems and investigate how the functionality of the ad-hoc node (within the vehicle) affects the quality of the traffic wireless information systems in ITS. The project covers the middle ground between VANET and collaborative data generation based on knowledge granularity (aggregation) [3]. It will investigate the designing, implementing and modelling the functionality of a condition identification algorithm for an intelligent node in ITS wireless information system that will be – at the same time – an active participant in the formation, routing and general network support of such systems and also act as in-car traffic information and real-time control generator and distributor [4].

Let us start by classify the discoverable road conditions based on 1) how many cars needed to discover certain road condition? Also, 2) what kind of parameters and variables needed to put it in its context (determine the road condition)? We tried to present the associated frequent asked questions, which may raises in terms of:

- 1) Broadcasting: How each node ensures that at least one node has received its message? How far should each message travel form the generated node (distance / number of hubs / time)?
- 2) Redundancy: Can we avoid sending the same message from multiple nodes? Who will stop forwarding the message?
- 3) Collecting data: (Granulation) what is the optimal number for each condition? (Percentage: number of cars with specific condition vs. total number of nearby cars). Which car will announces the warning message of the discovered road condition?
- 4) The main research objectives are to design the algorithms' functionality and to implement a model of the network with the required node features, which it is anticipated, will form the basis of a future real-life case study implementation.

2. Traffic Condition Detection Algorithm (TCDA)

Some road conditions can either be derived (assessed) from the activity of the individual cars' electronic helpers like ESP or ABS, or alternatively, sensors embedded in the individual vehicle may provide this information. The summary in the Table 1 is summary of the most common road situations with the causes of those situations considered as Non-conclusive Individual Car Sensed Data:

Table 1: Possible Road Situations (examples)	
Individual Car sensors data	Possible Reasons
Windscreen Wipers (goes ON)	Rain / cleaning / By Accident
ABS Control (Slippery Road)	Snow / Oil spot / bad tires / Bad driving
Fog light (ON).	Fog / By Accident / Since yesterday
Movement Speed (Slow).	Traffic Jam / Driver using the phone/radio
Reduce Speed (Unexpected)	Hazard Ahead/saw a friend or interesting place

But if we can share this data among all nearby cars, by Combining Individual Car Sensed Data (Table 2), the result will be:

Table 2: Certain Road Situation (examples)		
Individual Car sensors data	Optimum Num	The Reason
Wipers	10% of cars ON	Rainy
ABS Control.	5 cars = ON	Slippery (snow)
Slippery Spot.	2 cars	Slippery spot
Fog light.	50% cars ON	Fog
Movement Speed	60%Slow/Stop	Traffic Jam
Reduce Speed	5 cars in 1sec	Hazard Ahead

By comparing the two tables, you will notice that each case in the first table could happen because of many reasons, which make it non-conclusive piece of information. But in the second table, if we know the number of neighbouring cars that got the same situation (the numbers quoted in the table are representative rather than conclusive for the condition and represent a matter of future investigation in real-life experiments), we will be certain about the reason for that situation. This mechanism transfers the non-conclusive individual car sensed data into very important (conclusive) data to describe the surrounding road conditions. We should notice that the optimum number in the above table should be predefined and updatable by the system itself.

2.1. Algorithm Features

The algorithm is very flexible and has several variable parameters, which influence the final outcome, and this paper presents our conclusions in determining the optimal set of values:

- Using Variable Conditions Search Limitation (CSL): Control the searching area for any situation by using selection of parameters (number of hops from source, certain timeout, and/or distance from source).
- Multi-zones detection: in case of more than one zone, it can Detect each situation zone borders separately (even if they are overlapped). Then report them in one or multiple warning messages.
- Delay for data collection: Random time slots delay used before forwarding the received messages.
- Infrastructure less system.

2.2. Algorithm Description

Input:

Road situation detection messages (SDM) received. It generated by any node (I called it Active Node) who senses any road problem or certain road situation, each has at least the following information: Message id, Two lists of nodes and its positions: active nodes and non-active nodes.

Initialize:

$i \leftarrow \{0 \dots \text{number of nodes} - 1\}$
 $B_i \leftarrow \text{neighbor set of current node } N_i$
 $AN_i \leftarrow \text{detected Active Nodes IDs set by } N_i$
 $NAN_i \leftarrow \text{detected Non Active Nodes IDs set by } N_i$
 $SDM \leftarrow \text{road situation Detection Message}$
 $SWM \leftarrow \text{situation Warning Message}$

Event: new situation has been detected in the current node N_i

Add the current node ID to the local AN_i {if $AN_i \neq \emptyset$ }

Generate $SDM \leftarrow \{N_i(id), AN_i, NAN_i\}$

Forward SDM via 802.11

Event: new SDM_i message has been arrive at the current node N_i

extract from SDM_i data sets: $SDM_i(id), AN_i, NAN_i$;

if SDM_i is redundant **then**

discard SDM_i ;

else

$NAN_i \leftarrow NAN_i \cup NAN_i$

$AN_i \leftarrow AN_i \cup AN_i$ // update local lists of known AN & NAN

if $\frac{\text{length}\{AN_i\}}{\text{length}\{AN_i\} + \text{length}\{AN_i\}} \geq \text{optimum Number}$ **then**

// all data is Known

Calculate Zone // identifier zone situation by using AN list (NodeID and Position)

Generate $SWM \leftarrow \{N_i(id), S_i, zone, AN_i\}$ // generate new warning Message

Broadcast SWM via 802.11

else-if (distance between fairest two nodes in $AN_i \geq \text{Optimum number}$) **OR** (timeout) **then**

Generate $SDM_i \leftarrow \{N_i(id), AN_i, NAN_i\}$ //

generate new warning Message

Forward SDM_i via 802.11

else

Wait tolerant-time // to receive and collect

more data to broadcast all in one message.

update AN_i, NAN_i // update lists based on known AN & NAN

Generate $SDM_i \leftarrow \{N_i(id), AN_i, NAN_i\}$

//generate new Discovery Message

Forward SDM_i via 802.11

end-if

end-if

Event: new SDM_i situation has been received in the current node N_i

update AN_i, NAN_i // update lists based on known AN & NAN

Calculate Zone // identifier zone situation by using
AN list (NodeID and Position)

Generate $SDM_i \leftarrow \{N_i(id), AN_i, NAN_i\}$ //generate new
warning Message

Forward SDM_i via 802.11

3. Results

A model of an Adhoc network was set up to simulate the VANET using NS2 to generate traffic. Chosen parameters to simulate real life scenarios were assigned. The scenarios we

have used to test our algorithm were extracted from a real data monitoring system in Nottingham traffic control centre. This support increases the reality of our scenarios in all aspects (real road map, cars movement, road density, cars speed, etc). We should point out here that, the factor of having natural obstructs (building, Walls, Mounts ...) is not considered in our test model because of the difficulty of implementing or representing these factors in NS2 models. Ignoring these natural obstacles will boost up somewhat the number of dropped messages (interference) in our system, but – by percentage – we assume that this will not affect our results because of the minor effect and the iteration techniques that we used to calculate our results.

The simulation results are compiled on the basis of average results of running the simulation 10 times per each assumption (each Delay Time \times each max Hops \times each Active node number \times each total number of all nodes) with complete random selection of active nodes. The movement patterns we tested are two cases: first case, complete random generation in each round, second case, the same movement pattern for all rounds with the same delay and number of hops.

3.1. Results Aggregation

The simulation has been repeated 32,000 times. To help improve the aggregation and visualisation of the data and present them in dynamic form, a tool has been developed for that purpose. The tool can help in discovering the trend for any of the collected data (e.g.: number of exchange messages, max seen active nodes...) based on the number of hops and delay time. This approach makes us able to predict the optimum number of hops with the best delay time. We use it to establish the optimum parameters for the best performance of the algorithm (e.g.: reducing the number of exchanged message over the most suitable delay time with the maximum number of recognized active nodes and the maximum number of non-active nodes).

3.2. Results Analysis

As we are looking for the optimum number of hops to discover the whole local area and, at the same time, the optimum Delay time each node should use before resending any message, we analyse all the available data from the simulation with these two parameters (Num of Hops & Delay Time) as variables separately.

The number of exchange messages needed, message exchange time and number of recognized nodes (AN/NAN) are used as indicators for the best results and are sufficient to detect any Traffic Condition. The results for each are considered in the following:

Number of Exchanged Messages Per Node

Those Figures (1a, 1b, 1c) show the number of sent and received messages at each node and indicates how noisy the system is. It also gives an indication of the optimal value for the number of hops parameter

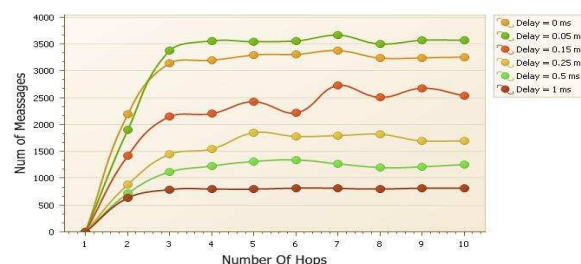


Figure 1a: Number of Received Messages /Node

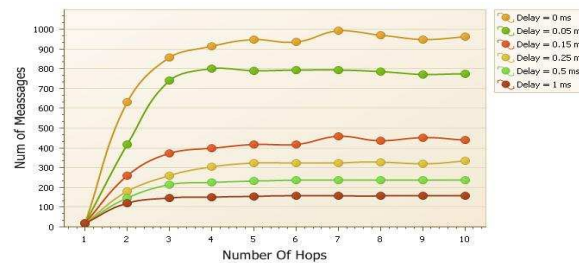


Figure 1b: Number of Sent Messages / Node

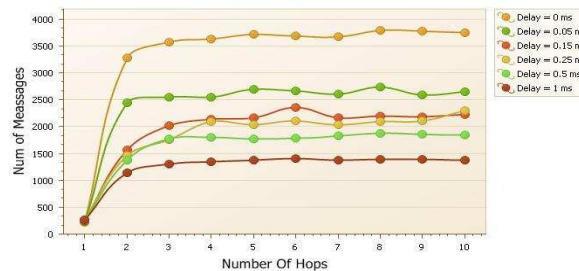


Figure 1c: Number of Discarded Messages/Node

Message Exchange Time Needed to Discover the area: The Figures (2a,2b) shows the total time needed for the algorithm to finish as a function of the number of hops parameter. Choosing the shortest total time needed to exchange all messages to detect a certain situation is important for the speed of detection and also for the timeout required before re-initiating the discovery sequence.

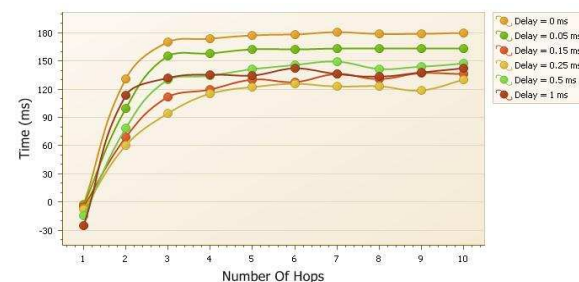


Figure 2a: Message Exchange Time (Total Time)

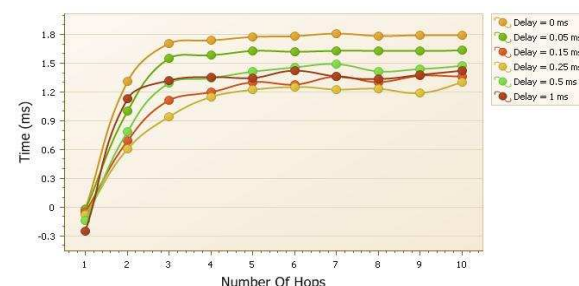


Figure 2b: Message Exchange Time (Time/Node)

Number of recognized nodes (AN / NAN)

Knowing the ratio between AN and NAN (or simply their numbers) is crucial to detect if the situation is present or not. The results of the experiments presented in the Figures (3a, 3b, 3c, 3d) show the number of recognised nodes as a function of the number of hops parameter.

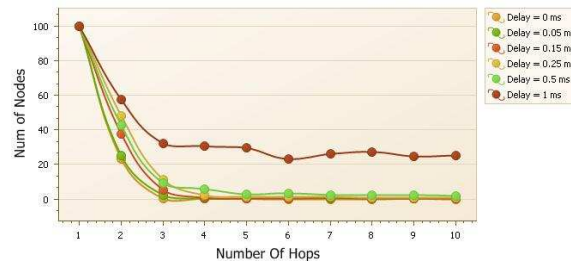


Figure 3a: Number of Nodes Saw up to 50% of AN

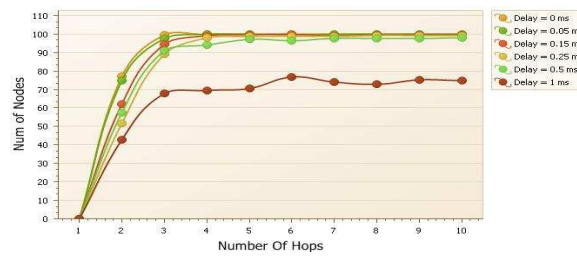


Figure 3b: Number of Nodes Saw more than 50% of AN

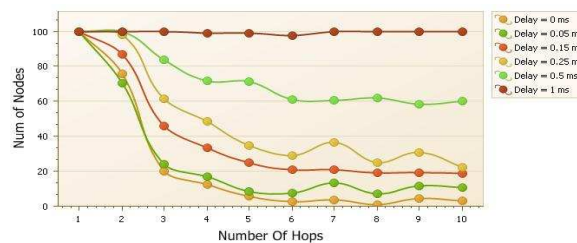


Figure 3c: Number of Nodes Saw up to 50% of NAN

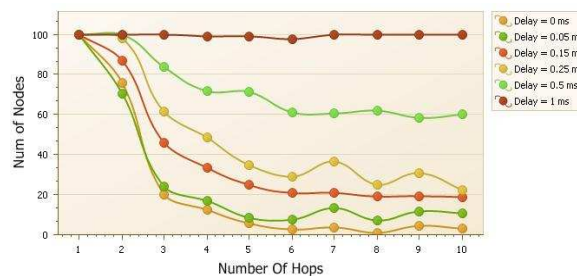


Figure 3d: Number of Nodes Saw more than 50% of NAN

3.3. Results Outcome

This study attempts to identify the optimum value for two algorithm parameters; number of hops and delay time. The assumption is that different situations are detected by different numbers of recognized AN/NAN (e.g: situation is rainy if 33% of nodes are AN, or a slippery spot can be detected by 3 AN regardless the number of NAN).

Analysis of the graphs presented indicates that there is no fixed optimum number either for delay time or number of hops. Consequently a range of numbers for these two parameters must be considered dependant on the detection cases. Based on these assumptions, we are looking for the best results that can recognize from 50% up to 100% of active nodes, which will be enough to cover all cases.

In the graphs presented the point of saturation i.e. where an increase in the value of the investigated parameter gives relatively small improvement in the quantity of

sent/received/discarded messages. The results show clearly that using **from 3 up to 5 hops** is optimum to detect any Traffic Condition if we consider the mentioned indicators.

The results show that the greatest delay time will reduce the number of exchanged messages, but will increase the total time needed to recognize the biggest possible number of AN/NAN. This is a difficult compromise between Time and noise, though a figure **between 0.01 and 0.1 second** seems to be indicated.

4. Conclusion & Future Work

An infrastructure-less vehicle-to-vehicle communication system in terms of data sharing and collaborative generation of information, as well as the characterized particular vehicular networks (moved nodes, road constrains mobility and variable communication conditions) is hot issue and research challenge for academics. Several dissemination protocols were proposed in research works. They could be sorted into two classes: (i) protocols for infotainment services (e.g. advertisement applications) that have constraints related to the bandwidth, and (ii) protocols for emergency services (e.g. road safety services) that have end-to-end delay and delivery ratio constraints. Also, Vehicular networks can be considered as the portal of many services, ranging from safety to traffic information and location based services (LBS). These services generally require efficient routing and dissemination protocols.

The proposed TCDA is a highly efficient protocol compared to pure flooding – the only algorithm reported so far capable of discovering reliably the information on an ad-hoc basis. Also, it has been proved that the algorithm can discover traffic conditions within certain areas using both dynamic variable search limitations and an intelligent routing mechanism. Optimal values for recommended number of hops and delay time have been identified and reported. It is clear that tomorrow's driving assistance systems can go far beyond their present capabilities by implementing co-operation and information exchange in order to collectively and cooperatively perceive the driving environment. Making decisions dependent on the environment can serve car drivers, ITS, environment and people more generally. This paper demonstrates a way of achieving this goal and paves the way for new and improved algorithms which to use car-to-car communication for traffic context identification. In this context the algorithm itself can be improved by identifying dynamically the boundary conditions as well as dynamic change of the traffic conditions for identification and employment of dynamic parameter restrictions.

References

1. Caveney, D. (2009) Cooperative Vehicular Safety Applications, in VANET: Vehicular Applications and Inter-Networking Technologies (eds H. Hartenstein and K. P. Laberteaux), John Wiley & Sons, Ltd, Chichester, UK. doi: 10.1002/9780470740637.
2. Eichler, S., C. Schroth, T. Koscht, and M. Strassbergert, 2006. Strategies for context-adaptive message dissemination in vehicular Ad hoc networks. Proceedings of the 3rd Annual International Conference on Mobile and Ubiquitous Systems, Jul. 17-21, IEEE Computer Society, San Jose, California, USA, pp. 1-9.
3. JA Fax, RM Murray, "Information flow and cooperative control of vehicle formations - IEEE Transactions on Automatic Control, 2004.
4. J. Jetcheva, Y. Hu, D. Maltz, and D. Johnson. A simple protocol for multicast and broadcast in mobile ad hoc networks. Internet Draft: draft-ietf-manetsimple-mbcas-01.txt, July 2001.
5. Khadraoui, D., Sukuvaara, T., "Wireless traffic service communication platform for cars", Information Infrastructure Symposium, 2009. GIIS '09, Global, pp. 1-7, 23-26 June 2009.

6. L. Krishnamachari, D. Estrin, and S. Wicker. The impact of data aggregation in wireless sensor networks. In ICDCSW '02 Proceedings of the International Conference on Distributed Computing Systems, pp. 575-578, 2002.
7. L. Zhou, G. Cui, H. Liu, D. Luo and Z. Wu, 2010. NPPB: A Broadcast Scheme in Dense VANETs. Information Technology Journal, Vol. 9, pp. 247-256.
8. Menouar H., F. Filali and L. Massimiliano, 2006. A Survey and qualitative analysis of MAC protocols for vehicular Ad Hoc network. IEEE Wireless Commun., Vol. 13, pp. 5-30.
9. Ni, S.Y., Y.T. Chen and J. Sheu, 1999. The broadcast storm problem in a mobile ad hoc network. Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking, August 1999, Seattle, Washington, pp. 151-162.
10. Thomas, M.; Peytchev E.; Al-Dabass D.; "Auto-sensing and distribution of traffic information in vehicular ad hoc networks", International Journal of Simulation, January 2004, Vol. 5(3), pp. 59-63, ISSN: 1473-804X.
11. Tseng, Y.C., S.Y. Ni and E.Y. Shih, 2003. Adaptive approaches to relieving broadcast storms in a wireless multihop mobile Ad hoc network. IEEE Trans. Comput., Vol. 52, pp. 545-557.
12. Wisitpongphan, N., O.K. Tonguz, J.S. Parikh, P. Mudalige, F. Bai and V. Sadekar, 2007. Broadcast storm mitigation techniques in vehicular Ad Hoc networks. IEEE Wireless Commun., Vol. 14, pp. 84-94.

Traffic Message Delivery Broadcast Protocol in Vehicular Ad Hoc Networks

Yueyue Li¹, Evtim Peytchev²

Nottingham Trent University, Clifton Lane, Nottingham, NG11 8NS, United Kingdom

¹yueyue.li@ntu.ac.uk, ²evtim.peytchev@ntu.ac.uk

Abstract: Traffic problems in the field of Intelligent Transport System (ITS) have always been an attraction in the researchers' eyes all over the world. To reduce traffic congestions, to save travel time, to decrease traffic accidents and to provide demanding information exchanges have become challenges of today and the future. Current research works focus on applying Car-to-Car (C2C) and Car-to-Infrastructure (C2I) approaches in infrastructure-less and flexible ad hoc networks environment. The routing problem has always been one of the most difficult problems in such dynamic environment network. This paper presents a novel, designed for routing purposes, traffic routing algorithm (TMDA) for a novel VANET architecture. The algorithm with the inclusion of urban traffic related routing information has been designed to be deployed in vehicles, e.g. cars and buses and aims to provide proper strategies for the utilization of travel information available in many of the vehicles traversing urban networks. The research investigates and compares communication performance of the communication system under TMDA and the other existing ad-hoc routing protocol (e.g. AODV) by a set of experiments with the NS-2 simulator. According to simulation-based performance evaluation, the proposed algorithm, TMDA, provides higher efficiency and reliability than a popular used broadcasting method for data dissemination.

Keywords: ITS, C2C/C2I, VANET, Routing Algorithm, NS-2 Simulation.

1. Introduction

In recent years, many more projects emerge in the field of Intelligent Transport System (ITS) because of the increasing importance of solving traffic problems, such as avoiding traffic jam and fast accident notifications etc. Fast and reliable real-time traffic information is irreplaceable tool to build safe and efficient traffic environment. To achieve this goal, traffic objects should cooperate with each other by using Car-to-Infrastructure (C2I) and Car-to-Car (C2C) communication approaches, as the communication of information is the biggest unutilised fully factor in ITS.

Typical examples adding weight to this concept are C2X communications investigated in the following projects of the 6th EU Framework Programme for Research and Technological Development [3]: 60 million EU CVIS (Cooperative Vehicle-Infrastructure Systems) Integrated Project [1], targeting mobile traffic participants to provide wise interactions between mobiles and transport infrastructures for road safety; COOPER (CO-OPERative SystEms for Intelligent Road) project [2], aiming at cooperative traffic management by exchanging real-time traffic information among travellers and fixed roadside system to finally enhance road safety on motorways.

In this paper, a novel Vehicle Ad-hoc Network (VANET) architecture for city traffic communications is introduced. This framework will create an opportunity for investigation of the benefits of car-based acquisition and dissemination of traffic information as well as generation and distributed implementation of traffic control. To investigate communication performance of such communication system we usually look at the routing protocols, which are basic necessity for having more reliable message transmissions from source nodes to destination nodes [9] for VANETs. This paper highlights implementations of innovative and feasible routing strategies model based on pre-existing node travel information. Particularly, a new Traffic Message Delivery Algorithm (TMDA) is created for the proposed model. The algorithm is based on broadcasting methods and covers a series of intelligent transmission

mechanisms. For example, the inclusion of traffic route information in the algorithm is able to provide routing plans for message forwarding and message receiving.

Compared with real test-beds, simulations can save large expenses to construct a model and allow components to execute repeatable tests in diverse targeting scenarios. Therefore, the majority of the researchers, nowadays, consider simulation technologies are indispensable [4]. This paper discusses essential simulation issues in details and displays results for investigations of the new routing algorithm in the proposed VANET architecture.

2. Related Work

Presently, a plethora of routing protocols is designed to adapt flexible and dynamic ad hoc networks. This paper will only concentrate on those studies being directly related to the proposed techniques and protocols.

Broadcasting is a basic method used in ad hoc networks. The simplest and earliest broadcasting technique is flooding methods, as described in [11]. Each mobile node, which receives the packet for the first time, periodically broadcasts or rebroadcasts the packet to all neighbours; otherwise, the receiver will discard the packet due to redundant operations.

The main problem of the simple flooding, also known as blind flooding [5], is the high amount of redundant broadcasting messages. This is referred as broadcast storm. To solve the problem, a few of solutions are proposed. For example, an IEEE802.11-based protocol named urban multihop broadcast (UMB) is designed in [8] to minimize the broadcast storm by allowing the farthest vehicles to receive and forward data and inform other nodes between original senders and itself. Meanwhile, it uses acknowledgment messages (ACK) to guarantee high reliability of packet delivery.

In the traffic area, diverse and changeable communication demands and traffic problems can occur at any time and every day. For these reasons, maximum and optimum information are expected to be included in communication protocols by many research and projects. Some researchers have proposed algorithms with the inclusion of particular traffic information, for example, the inclusion of the acknowledgments into the periodic beacons for high reliability [10] and the inclusion of vehicles' status and surrounding information in [6] etc.

3. The Novel VANET Architecture

The paper introduces a VANET architecture that, based on C2I/C2C communications, involves spontaneous wireless communications occurring within a group of wireless mobile nodes (Figure 1). The architecture integrates features of traditional ad hoc networking technologies and VANET technologies, being used in standalone mode or cooperative connections to the larger Internet [4].

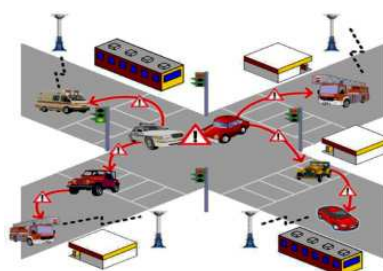


Figure 1: Novel VANET architecture
Ref.: <http://www.ics.uci.edu/~keldefra/manet.htm>

Being different from traditional ad hoc networks, the communication system utilizes vehicles for routing purposes via the inclusion of traffic route information. It recognises three types of ad hoc nodes – mobile, semi-mobile and static ad hoc nodes. Each type of node is expected to exert their separate functionality when communications occur. Mobile nodes, such as cars, are defined as traditional ad hoc nodes without pre-conceived route with functions of routing and transmitting messages. Alternatively, bus-nodes, considered as semi-mobile nodes – having predetermined route onto which they are currently travelling, integrate routing, transmitting and gateway altogether to provide a possibility of interconnection among ad hoc networks and communications among ad hoc networks and wired networks communication. They work as common ad hoc network only when occasional events occur, such as disconnections at night or traffic jams beyond bus-lanes etc. In those cases, bus-nodes exert their advantages about structured routes and stable running timetables to offset environment shortcomings. As far as static ad hoc nodes are concerned, they will cooperate with other two types of nodes to provide more reliable and detailed information. They are assumed as Internet-enable nodes to provide access to the other type of networks. Notably, they have closer relationship with bus-nodes rather than cars because the location of these static nodes is mostly next to the bus-lanes.

4. The New Traffic Message Delivery Algorithm (TMDA)

Traffic Message Delivery Algorithm (TMDA) is a novel traffic routing algorithm designed for improving communication performance of a particular VANET network described in section 3. In TMDA, the approaches of the message delivery vary relying on the features of nodes, including the type, position, direction of each communication vehicle, at that moment.

Compared with common lanes, the routes mainly used by bus-nodes have fewer disconnections even in a sparse network because bus-nodes are appearing on the routes regularly. Once static nodes (e.g. Roadside units or speed 0 mobiles with Internet connections) cooperate with them, the information could be quickly shared in the limited local areas or even the whole city area.

One of highlights in TMDA is the inclusion of I-Routes, which is simplified as bus-lanes in this research. Bus-lanes exist with fixed directions, locations and even required speeds. Thus, they greatly help to guide the whereabouts of messages carried by vehicles.

Algorithm 1 exhibits detailed receiving processes of TMDA. Generally, if messages reach I-Routes, they will be faster forwarded following the pre-configured directions of the I-Routes; otherwise, they are based on developed broadcasting strategies only. Procedures are relatively simple for sending the message that nodes carry on periodic broadcasting via IEEE 802.11 within a certain expiry; whereas more considerations occur in terms of receiving a message. Further details of message receives are explained by the following pseudo-codes.

Algorithm 1 Pseudo-code of Message Receiving

```

1 Event: the message has been received
2 if msg_id is not in check_list then
3   receives the message
4 else
5   discard the message

6 Event: the message received from NB or
7 if R = src then

```

```

8  discard the message;
9  else
10   if  $R = dst$  then
11     inform others to stop broadcasting;
12   else
13     if  $P_s$  is on I-Routes then
14       if  $P_r$  is on I-Routes then
15         when  $T_c = T_{d1}$ , farthest nb forward message;
16         inform others between  $\langle S, R \rangle$  to stop broadcast;
17         message is stored longer in this node  $R$ ;
18       else
19         if  $D_r = D_s$  then
20           when  $T_c = T_{d2}$ , farthest NB forward message;
21         else
22           when  $T_c = T_{d3}$ , farthest NB forward message;
23       else
24         if  $P_r$  is on I-Routes then
25           when  $T_c = T_{d1}$ , farthest nb forward message;
26           inform others between  $\langle S, R \rangle$  to stop broadcast;
27           message is stored longer in this node  $R$ ;
28         else
29           when  $T_c = T_{d1}$ , farthest nb forward message;

```

where:

msg_id represents a traffic event and it is recorded in a receiver's check_list at the first receiving per broadcast;

R indicates current receiver and the situations are source (src), destination (dst) or intermediate nodes;

P presents the position of nodes. P_s and P_r are the position of the last sender and current receiver respectively;

D implies the direction of nodes. D_r and D_s are the direction of the last sender and current receiver respectively;

T means the forwarding time. Indicates current time. Particularly, TMDA designs waiting intervals for nodes in different conditions; thus, T_{d1} , T_{d2} , T_{d3} consist of both receiving time and waiting time and the order is $T_{d1} < T_{d2} < T_{d3}$.

5. Simulation Issues

NS-2 is selected as a well-suitable simulation tool in this paper. It uses Tcl (Tool Command Language) to organize script files for setting up traffic patterns such as scenarios and movements and also communication patterns, e.g. transmission issues.

5.1. A City Scenario

In terms of traffic patterns, the focus at this stage is #-shaped city scenario (Figure 2).

#-Shaped city scenario (Figure 2) – a medium scale network with possible traffic units consists of intersections, horizontal and vertical roads.

Nodes – The term density represents as the number of nodes over the network. This paper presents two dense levels (Figure 3), low and high respectively.

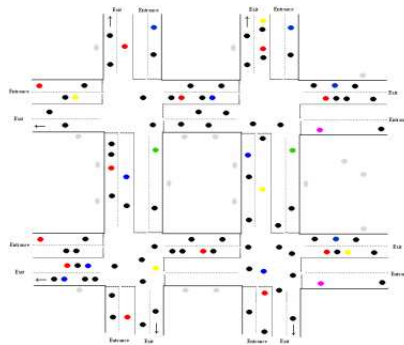


Figure 2: #-shaped traffic pattern

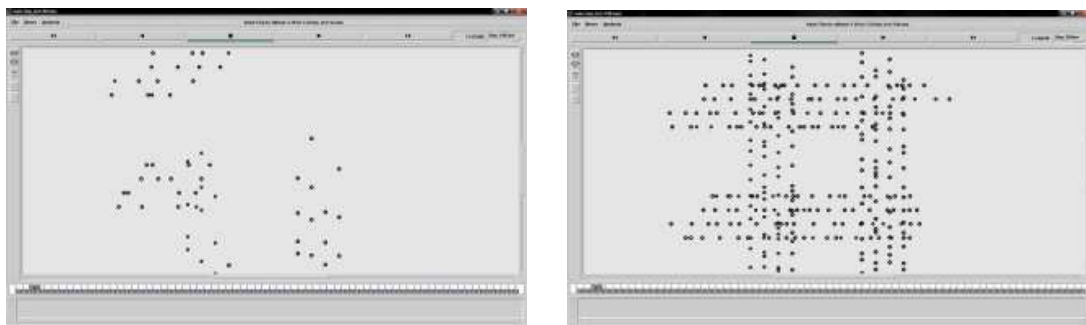


Figure 3: Simulation models for various densities of networks

I-Route – This is a term for a set of special routes integrated in our established ad hoc wireless mobile communication system. They are assumed as bus-lanes. Compared with other urban lanes, message transmissions on I-Route have higher priority and determined broadcasting directions. The overall aim of the inclusion of I-Routes in TMDA is to increase communication performances in aspects of Current simulation models adopt the following I-Route patterns drawn as two lines with arrows in Figure 4.

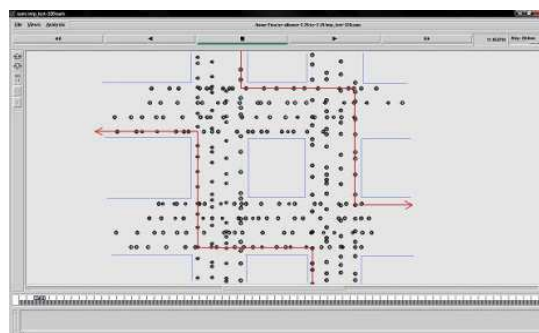


Figure 4: Simulation model with I-Routes

5.2. Transmissions

Following points are essentially to impact on the design of simulation models.

Range – The simulation model sets transmission activities on the IEEE 802.11 standard. Our simulation models in this paper specify Two Ray Ground propagation model with the transmission range about 150 meters.

Distance & Speed – In our designs, the nodes are distributed following the shape of urban lanes. The distance between two nodes is randomly chosen based on the minimum 150

meters; and the speed of nodes in urban scenario is also randomly generated but no more than 20 m/s.

Message – Message contains three elements: message size, message id and attached information, such as source node (SRC), destination node (DST), current sender (Sc), the position of senders (Xc, Yc), the speed of senders (SPc), the direction of senders (DIRc), the message expiry (EXP) and current timestamp (Tc).

6. Results Evaluation and Analysis

6.1. Network Communication Performance Metrics

End-to-End Delay Time (EDT) – It refers to the duration of a message sent from source to destination over the network [7]. The equation (1) solves the average delays (ΔEDT) by using the sum of single delays ($\Sigma(EDT)$) and the number of tests (n).

$$\Delta EDT = \Sigma (EDT) / n \quad (1)$$

Message Delivery Ratio (MDR) – It represents a ratio of successful message deliveries. The final evaluation of this paper will follow the results obtained via equation (2) which shows the average value of the testing delivery ratios.

$$\Delta MDR = \Sigma (MDR) / n \quad (2)$$

6.2. Results in Various Dense Networks

On the basis of urban scenario shape (section 5.1), we consider totally approximately 4000 simulation runs for two kinds of density networks. In this paper, the word ‘dense’ means overall 320 vehicles in the whole network area (6000×6000m) and the term ‘sparse’ indicates to about 65 vehicles in such network.

Shown in Figure 6, 7, 8 and 9, TMDA provides better communication performances than the evaluated current ad hoc routing protocol (AODV). The further analysis will be divided into two parts: 1) dense network scenario; 2) sparse network scenario.

6.2.1. High Density

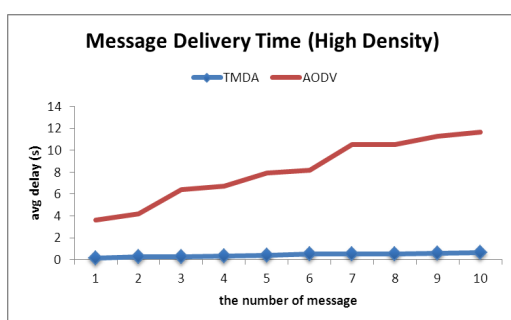


Figure 6: Delays in the high & medium density of networks

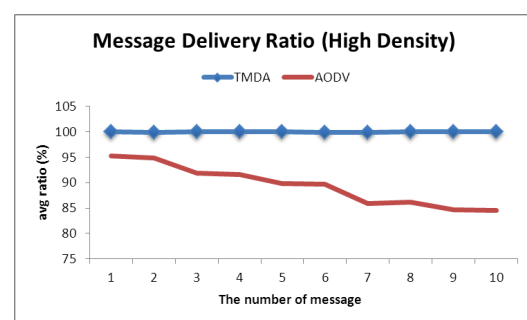


Figure 7: Rates in the high & medium density of networks

Figure 6 represents the average EDT and Figure 7 shows the trend of MDR in a dense network. TMDA exhibits smaller and more stable EDT from 1 message to 10 messages per testing time, reflecting on the below lines in Figure 6 and higher MDR from the above lines in Figure 7 than those obtained from AODV protocols. Conversely, AODV shows that the trend of the average EDT in the network goes up accompanied with increases of the message number shown in Figure 6; as well as the ratio displays as decreasing status in Figure 7.

One of drawbacks inherited from AODV is the broadcast storm which is also considered as a major reason of packet loss. However, TMDA adopts delay strategies to reduce broadcast storm and the results prove that the packet loss is relatively less prominent.

6.2.2. Low Density

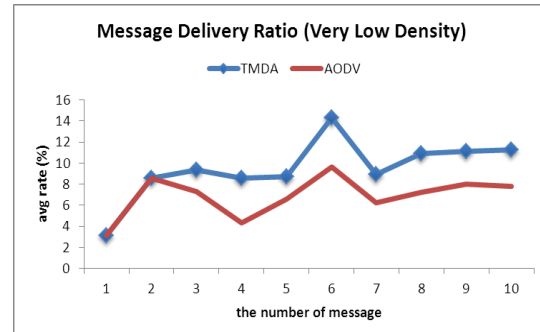
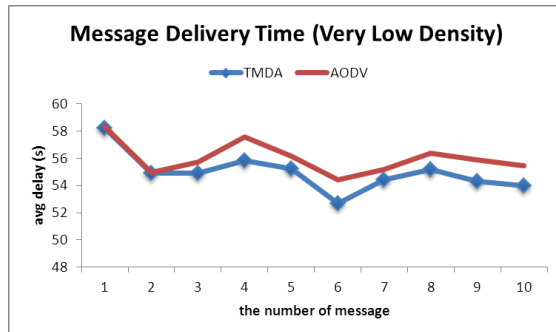


Figure 8: Delays in the low density networks **Figure 9:** Rates in the low density networks

Figure 8 and Figure 9 display average EDT and MDR in low density networks respectively. TMDA provides better results than those of AODV. For example, EDT lines of TMDA in both networks are lower than those of AODV with smaller average delays. Meanwhile, the above MDR lines which represent higher successful packet deliveries are from TMDA.

The reason is that TMDA allows senders to continually broadcast the message within the expiry unless they receive a redundant message or they receive an instruction included in the message to stop broadcasting. The approach saves the time for senders to wait the response and also each sender needs not to keep a list to record paths for replying packets. Moreover, TMDA contains I-Route information, which helps to improve the ratio of message deliveries, particularly in sparse networks. In our experiments, we set maximum expiry for non-emergency messages to be 60 seconds. That is, any delay time more than 60 seconds will be regarded as final packet loss.

Generally, the average message delivery time, seen in Figure 8, are smaller by using TMDA from 1 message to 10 messages.

7. Conclusion and Future Work

This paper presents the comparisons of communication performance by using different routing protocols in a novel VANET architecture. The newly created algorithm TMDA, not only adopts principles based on existing broadcasting algorithms but also incorporates urban traffic route information into the algorithm, utilizing the concept of 'I-Route' available in vehicles. The aim of these new routing strategies is to alleviate the impact of the problems caused by previous routing protocols and also best service for the particular implementation background.

The future work will concentrate on applying the algorithm in a real city scenario (e.g. Nottingham city) to further investigate above results of simulations. Meanwhile, static nodes are considered to be integrated into the architecture for collaboration studies.

References

1. CIVS (Cooperative Vehicle-Infrastructure Systems), Available: <http://www.cvisproject.org> [Feb, 2012].
2. COOPER (CO-OPERative SystEms for Intelligent Road), Available: <http://www.coopers-ip.eu> [Feb, 2012].
3. European Commission Community Research-CORDIS. The Sixth Framework Programme in brief. 2002.
4. Imrich Chlamtac et al., Mobile ad hoc networking: imperatives and challenges, SCIENCE@DIRECT, 2003.
5. Stojmenovic and J. Wu, "Broadcasting and Activity Scheduling in Ad Hoc Networks," Mobile Ad Hoc Networking, S. Basagni, M.Conti, S. Giordano, and I. Stojmenovic, Eds., pp. 205-229, IEEE Press, 2004.
6. J. Fukumoto, Sirokane, Y. Ishikawa, Wada, N. Tomotaka, K. Ohtsuki, H. Okada, Analytic method for real-time traffic problems in VANET by using Contents Oriented Communications, 7th International Conference on ITS, 6-8 June 2007.
7. L. Feeney, B. Ahlgren, A. Westerlund, Spontaneous networking: an application-oriented approach to ad hoc networking, IEEE Communications Magazine, 2001.
8. LI Li-Jun, LIU Hong-Fei, YANG Zu-Yuan, GE Li-Jia, HUANG Xi-Yue, Broadcasting Methods in Vehicular Ad Hoc Networks, *Journal of Software*, Vol.21, No.7, July 2010, pp.1620–1634 (in Chinese with English abstract).
9. PIORKOWSKI, M., Collaborative *Transportation Systems*, 2010 IEEE Wireless Communication and Networking Conference, pp. 1-6.
- 10 .Ros, F.J.; Ruiz, P.M.; Stojmenovic, I., "Acknowledgment-Based Broadcast Protocol for Reliable and Efficient Data Dissemination in Vehicular Ad Hoc Networks," *IEEE Transactions on Mobile Computing*, Vol. 11, No. 1, pp. 33-46, Jan. 2012.
11. T. Camp and B. Williams. Comparison of broadcasting techniques for mobile ad hoc networks. In *Proceedings of The Third ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC 2002)*, Lausanne, Switzerland, Jun 2002.

Mobile Network Architecture for Application Support in Remote Location

Alroqi Yasser¹, Evtim Peytchev²

School of Science and Technology, Nottingham, UK

¹yasser.alroqi2010@ntu.ac.uk, ²evtim.peytchev@ntu.ac.uk

Abstract: A mobile ad hoc network is a self-organizing and rapidly deployable network in which neither wired backbone nor a centralized control exists. It would be a of great advantage for the ad-hoc networks if they can provide the opportunity for nodes to collaborate for delivering services to the network members. Intensive connectivity message exchange between mobile nodes to address the dynamism of mobile ad hoc networks (MANETs) and the emerging problems thereof have been the core research areas for MANETS. Previous studies showed the usefulness of MANETs in order of collaborative mobile nodes to achieve better performances, offer more services or solve problems.

These research efforts have led to the creation of a clear technical basis for dealing with the aforementioned problems regarding node connectivity in MANETs. While the general principals of such networks are well known, when it comes to specific types of applications the principles behind providing support for such applications are not well investigated. This is especially true for application that employs collaborative approach in achieving the intended results.

Keywords: Mobile Collaborative, Two Level Routing Protocols, Application Support in Remote Location.

1. Introduction

Wireless mobile ad hoc network (MANET) is a network that consists of mobile devices, and each mobile device communicates with others via radio channels without presence of fixed network infrastructure. Previous studies have shown the usefulness of MANETs in order of collaborative mobile nodes to achieve better performances, offer more services or solve problems, e.g. Coordination in common emergencies, military activities ,and search and rescue operations[11][10].

Defining the necessary computational characteristics of the node (network node computational power, data processing device rules, wireless connectivity etc.) in order to support the identified collaborative application functionality and the necessary software support for message exchange in cooperative applications and knowledge generation in distributed remote environment will help us to determine how can the communication ad-hoc network routing protocols be open to include specialised ad-hoc network collaborative application configuration and usage.

One useful possibility is using knowledge of location of nodes in the network to support each participant to determine site next to it. Coordinates and comparing location is the way that can define how proximity the distance between nodes. This information in turn can control the type of routing protocol which will be used either to be at the application level or the routing level and thereby reducing the amount of message overhead. Furthermore, this will increase the performance of the network and try to gain reliability and stability of the network.

This paper describes a newly started PhD project that will investigate reactive routing on several layers – general routing for MANETs which will allow nodes to communicate together for exchanging data and using each other's services in the general case, and on different level providing communication architecture, tools, and API protocol for services provision and discovery specifically tailored to distributed collaborative applications. This is especially

useful when we consider remote locations where the only services available reside within the nodes themselves. The project aims to investigate the possibility of gaining ad-hoc cooperation information generalisation, and determine how the functionality of the ad-hoc node affects the quality of the wireless traffic information systems when collaboration is involved. It will investigate the possibility of designing, implementing and modelling the architecture of such an intelligent ad-hoc network support that will be at the same time active participant in the formation, routing and specialised collaborative application network support for MANETs. Furthermore, the investigation will consider the use of two-level routing protocol defines by the applications to support the combination, virtualization, and aggregation.

Next two sections describe the motivation and contribution and related works. Sections 4 and 5 present the current situation, and the future work for this work. Section 6 presents the conclusion.

2. Motivation and Contribution

There are communication requirements in order to enable group work in ad hoc network setting. One of these requirements is Dynamic Network Architecture which force the groupware system to adapt itself to determine the distribution schema that are best suited for a particular application and collaboration scenario because dynamism of topology, performance of network and reliability[9].

From that follows the idea to improve the scalability and efficiency of the service discovery under the ad hoc network through low level routing protocol taking into account specifics of the geometry of the network.

The contribution is investigating and designing a generalization model of ad-hoc network level support for distributed collaborative applications on the basis of investigating and identifying the requirements of such applications towards the underlying connectivity model. Moreover, designing, building and evaluating an architecture prototype model of multi-layered connectivity ad-hoc networks support for collaborative application (case study – multi-layered approach for collaborative applications running in ad-hoc networks in remote locations or on a large scale urban network) and building a model for collaborative application communication support, identifying suitable simulation environment and designing a set of experiments for proof of concept testing for real life scenarios.

3. Related Work

The service means in the ad hoc network any applications or hardware objects that requested or utilised by users in the network. These services are generally divided (based on the type of protocol they utilise) into two groups: service discovery protocol (SD) and service advertisement protocol (SA). Firstly, SD protocols are “network protocols that allow automatic detecting devices and services offered by nodes on ad hoc network” [3]. Secondly, SA protocols refer to “network protocols that allow automatic attributes of services to be introduced to the network to access to all services easily [3].

The topology in ad hoc network is usually highly dynamic including frequent partitioning and merging of network components so, nodes should be able to define the location of services providers [2]. In addition, defining the current location of the service providers is important and some protocols have been proposed for service discovery and advertisement (SA/SD) [6]. Researchers looking into the differences between implementing the SD/SA on application layer and on routing layer based on the message exchange overhead have found that implementing in routing layer decreases the message exchange overhead

because they use only communication messages by piggybacking the SD/SA messages onto the routing messages.

Koodli and Perkins in [12] extended a reactive routing protocol namely Ad Hoc On Demand distance Vector protocol (AOVD) to provide the idea of routing layer support for service discovery. J. Antonio and Dante Airias in [5] have integrated the Ad hoc On-Demand Distance Vector (AODV) routing protocol with service discovery functionality and have experimentally compared it with NOM [3] (a pure application based service discovery protocol). Their findings show that the integrated protocol produces 30% to 50% less control overhead and has 2 to 7 times lower service acquisition latency than the application layer based protocol (depending on simulation parameters).

Comparing a reactive routing and service discovery protocol and a proactive routing and service discovery protocol, DSR and the Destination-Sequenced Distance Vector protocol (DSDV) were extended to provide service discovery functionality against Service Location Protocol (SLP) done by [12]. The extended DSR protocol proves to have the minimum messaging overhead among the three, with second best the extended DSDV protocol. DSDV is not the only proactive routing protocol extended with service discovery functionality. In [4] and [7] researchers have also extended the Optimized Link State Routing (OLSR) proactive routing protocol to support service discovery.

Comparing the proactive, reactive and hybrid integrated protocols one can see that the most energy efficient and also effective protocols are the hybrid ones [8]. Christopher N and George C. In [2] authors did a survey with regards to service advertisement, discovery, and selection schemes for MANETs, presenting the most representative approaches. Their analysis led to the identification of the need for autonomic service discovery protocols capable of flexibly adapting their operation (in terms of selected architecture, discovery mode, and disciplined values for parameters) to the actual context and service demand specific to the ad hoc network in which the protocol is used [1].

Using application programming interface is not new idea in the ad hoc network [13] implementing the HLMP (High Level MANT Protocol) which enhance mobile collaborative development with API. However, HLMP is proactive routing protocol and different from what we will plan to do in this work. We are focusing using reactive routing protocol.

4. Current Situation

For the time being this project is looking into the field of Mobile Ad hoc Network aiming to identify core communication functionality of the existing intelligent applications and to identify their requirements towards the underlying ad-hoc networks. In addition, the project will identify the differences between single node and collaborative nodes applications and generalisation of the communication requirements of the collaborative applications towards the underlying ad-hoc networks support. Moreover, a definition of the concept of two level routing protocols (generalisation – multi-layered will be designed and presented.

5. Future Work

The project will design the API for collaborative application communication support based on two level routing protocols. In addition the project will design a new cooperation model for new data and control generation based on two-level routing protocol – thus establishing rules for collaborative application design. As additional aim is identifying and deploying suitable simulation environment capable of simulating mobile wireless ad-hoc networks and scalable collaborative applications implementations.

Later stages of the project will deal with the design the simulation model for two-level routing ad-hoc network communication support leading to the design of the collaborative application generalisation communication model as part of the overall simulation model.

The project will evaluate the simulation models to identify and optimise the best parameters for the models on a limited scale using test data. In addition, the work will be extended to include using real-life (full scale) data in the simulation model with – traffic network topology in Nottingham embedded. Real-life use-case scenarios for modelling and evaluation through simulation will be identified. Finally, the project will conduct limit real-life trials – setting up hardware, connectivity technology etc. and running with limited number of devices real life experiments.

6. Conclusion

This paper describes the scientific foundations of a PhD project that will investigate the use of reactive routing protocols to get on demand application support, especially when we consider remote locations where the only services available reside within the nodes themselves. We plane to investigate the use of several layers routing protocol. First one allows nodes to communicate with together through the routing level. The second one provides architecture, tools and API protocol for service discovery. The main objective of the project is to improve the scalability and efficiency of the service discovery under the ad hoc network conditions through tow level routing protocol. Dynamic network architecture is a necessity and it will force the groupware systems to adapt itself to determine the distribution schema that is best suited for a particular application and collaboration scenario.

References

1. Arora, H.Z.A. et al., 2003. GS³: scalable self-configuration and self-healing in wireless sensor networks. *Computer Networks*, Vol. 43(4), pp. 459-480. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:GS+3+:+Scalable+Self-configuration+and+Self-healing+in+Wireless+Networks+*#0 [Accessed February 22, 2012].
2. C. N. Ververidis and G. C. Polyzos, 2008. Service Discovery for Mobile AD HOC Networks: A Survey of Issues and Techniques. *Communications*, pp. 30-45.
3. Doval, D., 2002. Nom: Resource location and discovery for ad hoc mobile networks. In *Proceedings of the First Annual Mediterranean Ad Hoc Networking Workshop*, Sardegna, Italy. Citeseer. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.20.3990&rep=rep1&type=pdf> [Accessed February 22, 2012].
4. Engelstad, P.E., 2006. Service Discovery Architectures for On-Demand Ad Hoc Networks. *Science*, Vol. 1, pp. 1-32.
5. Garcia-Macias, J.A. & Dante Arias Torres, 2005. Service Discovery in Mobile Ad-Hoc Networks, *Better at the Network Layer*, Processing.
6. Hosseini Seno, S.A., Budiarto, R. & Wan, T.-C., 2011. A routing layer-based hierarchical service advertisement and discovery for MANETs. *Ad Hoc Networks*, Vol. 9(3), pp. 355-367. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S157087051000082X> [Accessed January 25, 2012].
7. Jodra, J.L. et al., 2006. Service discovery mechanism over OLSR for mobile ad-hoc networks. *20th International Conference on Advanced Information Networking and Applications*, Vol. 1 (AINA'06), Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1620435>.
8. Li, L. & Lamont, L., 2005. A lightweight service discovery mechanism for mobile ad hoc pervasive environment using cross-layer design. In *Pervasive Computing and Communications Workshops*, 2005. PerCom 2005 Workshops. Third IEEE International Conference on. IEEE, pp. 55-59. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1392800> [Accessed February 22, 2012].

9. Meseguer Pallarès, R. et al., 2009. Building Real-World Ad-Hoc Networks to Support Mobile Collaborative Applications: Lessons Learned. Lecture Notes in Computer Science, Vol. 5784(Springer), pp.1-16.
10. Mousannif, H., Khalil, I. & Al Moatassime, H., 2011. Cooperation as a Service in VANETs. Journal of Universal Computer Science, Vol. 17(8), pp. 1202–1218. Available at: http://www.jucs.org/jucs_17_8/cooperation_as_a_service/jucs_17_08_1202_1218_mousannif.pdf [Accessed February 16, 2012].
11. Neyem, Andrés, 2008. Integrating Service-Oriented Mobile Units to Support Collaboration in Ad-hoc Scenarios. Computer, Vol. 14(1), pp. 88-122.
12. R. Koodli and C. E. Perkins, Service discovery in on-demand ad hoc networks. IETF Internet Draft, draft-koodli-manet-service-discovery-00.txt.
13. Rodríguez-covili, J., Ochoa, S.F. & Pino, J.A., 2010. HLMP API/: A Software Library to Support the Development of Mobile Collaborative Applications. Communication, pp. 479-484.

Traffic Control Strategies and Traffic Prediction in Wireless Traffic Information Systems

Qiming Zhang¹, Evitn Peytchev²

Nottingham Trent University - School of Science and Technology

Computing and Informatics Building, Clifton Lane, Nottingham, NG11 8NS, UK

¹qiming.zhang2006@my.ntu.ac.uk, ²evtim.peytchev@ntu.ac.uk

Key words: Traffic Simulation, Traffic Prediction, car-following models, Traffic Assignment.

Abstract:

This paper represents a roadmap for research in the area of traffic simulation systems and traffic prediction systems with the aim of designing, implementing and evaluating new distributed between the cars paradigm to traffic simulation. The research will look into existing traffic simulation models and evaluate their suitability for implementation in a distributed environment. On the basis of this investigation a new model incorporating every suitable feature of the models developed so far will be suggested, implemented and evaluated. It is envisaged that the proposed model will be valuable edition to various FP7 EU research projects.

1. Introduction

1.1. Traffic Control

Traffic control is also known as the traffic signal control or urban traffic control. It relies on the traffic polices or traffic signal control facilities directing the vehicles and pedestrians in a such a way as to minimise traffic delay and travel time. In order to achieve the purpose of reducing the traffic, ensuring the traffic safety and clearance, the traffic polices should execute restricting, regulating, leading and avoiding actions for the traffic flow through the traffic control facilities. Traffic control also involves controlling vehicles and pedestrians for some security reasons in some or all of the traffic sections. The transportation departments develop specific provisions for vehicles and pedestrians on the roads or other transport-related activities according to the law. The rules usually include traffic dispersion, traffic organization, traffic evacuation, prohibiting, restricting or indicating. These controls are temporary and citizens should comply with the notices.

1.2. Development

Traffic signal control originated in Britain in 1868 [1]. People used gas as a signal light on the London streets, which marked the beginning of the urban traffic signals. In 1926, US's town Chicago adopted a new program for traffic light control, that is, each intersection has only one traffic light. Since then, the traffic control technology and control algorithms have undergone rapid development and improvement, affecting traffic safety and effectiveness and also reducing the impact on the environment. With the development of modern information technology, electronic technology, automatic control technology and computer technology, traffic signals have changed from manual to automatic, from fixed cycle to variable cycle. And the system control mode changed from Isolated Intersection control and arterial wire control to areas-wide control, and then intelligent traffic systems (ITS) have been developed. Until the early 1980s, over 300 cities in the world have a traffic control centre with some sort of ITS implemented. The most representative of these control systems are the urban road traffic control systems TRANSYT and SCOOT by Transport Research Laboratory (TRL), and the SCATS system from Australia [2].

1.3. Traffic prediction

The traffic information prediction is an important part of the modern science of traffic and intelligent transportation systems. It is based on history or existing transportation factors and statistics, then uses the intelligent computational method to forecast the transport system state in the target area in the future.

The importance of traffic information prediction:

(i): Traffic information prediction can grasp the development of the local transport business and carrying trade. This is an important basis for development of transport development strategies and policies.

(ii): Its role is mainly reflected in the forecast of socio-economic, travel and other aspects. It is an important part of urban transport plan, that is, the traffic information prediction is the premise of the transport planning program.

(iii): Traffic information prediction is a technical support for traffic incident detecting, improving the forecasting capability of the traffic incident, reducing the negative impact of the event.

(iv): As an important information processing technology, traffic information prediction can not only constitute independent functional systems, but it is also an important part of many ITS systems. There is a relationship between traffic information prediction and ITS at below:

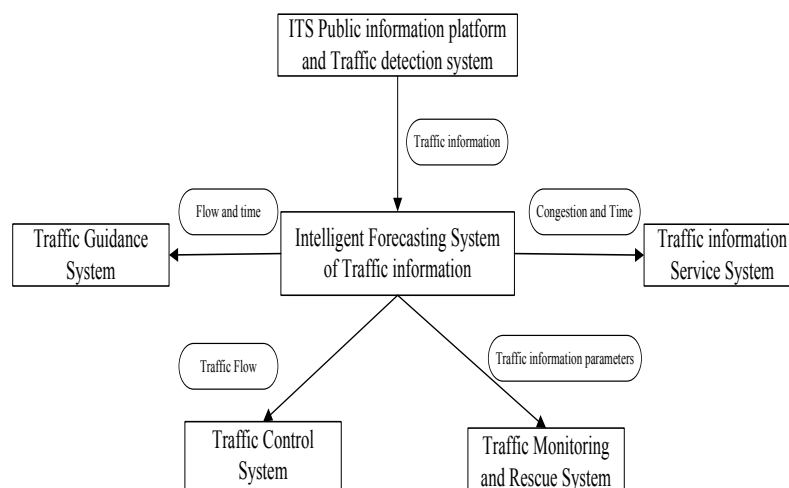


Figure 1.1

2. Traffic Information Prediction

2.1. Prediction Model

In the case of the current development of forecast, there are many traffic information prediction models and methods. It includes the Regression Analysis Method, Time Series Prediction, Grey Theory Prediction Model, Markov Predication Model, Neural Network, Kalman Filter and so on. However, each prediction method contains several prediction models. For example, when prediction model is based primarily on the traditional mathematical and physical methods, it should involve Time Series Prediction Model, Kalman Filter Model, Parameter Regression Model and Exponential Smoothing Model. If the prediction model uses modern science technologies as a research tool, the no-Parameter Regression Model, Warelet theory, Multi-fractal Method and Compound Forecasting Model of Neural Network are involved.

2.2. Classification of Prediction Models

The traffic forecast method reflects diversity and it was classified on the basis of the following six aspects [3].

2.2.1. Cycle Length

The forecast period could be a year, month, day, hour or even minutes for different forecast demand. The corresponding prediction method is divided into long-term, medium and short-term.

2.2.2. Characteristic

It mainly includes the quantitative prediction method and qualitative prediction method. If the characterization of the traffic information is a series data, the quantitative prediction should be taken. Otherwise, the qualitative prediction method will be used.

2.2.3. State

It is usually divided into static prediction and dynamic prediction. Static prediction does not contain the time changes, and predict the causal relationship of transport phenomena in the same period. Dynamic prediction need to consider the time changes, and according to the history and current situations of the transport phenomena, when people forecast the traffic trends in the future.

2.2.4. Traffic Kinetic Characteristics

This category includes deterministic prediction, chaotic prediction and stochastic prediction. The urban transport system is composed by people, vehicles, road, weather and other factors. Its essence is a complex non-linear system. In the different traffic conditions, traffic flow has these three kinetic behaviours.

2.2.5. Number of Prediction Models

When people forecast the traffic information, they would need to choose a single prediction or combination prediction. Single prediction means just use only one model to forecast the traffic information. And the combination prediction can use more than one model to forecast, so that it has the complementary advantage features.

2.2.6. Technology

It refers to conventional prediction and intelligent prediction. Many methods and models mentioned at above, conventional prediction usually includes Regression analysis Method and Time Series Method of Exponential Smoothing, Moving Average, Seasonal Coefficients and Box-Jenkins. Intelligent prediction comprises Grey Theory, Kalman Filter, Chaos Theory, Neural Network and multi- Agent prediction method [4].

2.3. Traffic Simulation

Traffic simulation applies the digital models to reflect the complex traffic phenomenon, and analysis the traffic situations. According to the detail's levels of the simulation model described in the transportation system, the traffic simulation is usually divided into microcosmic traffic simulation model and macroscopic traffic simulation model, even sub-micro or pure-micro and mesoscope (meso) will be used in some cases.

Microcosmic traffic simulation model has a high description of the traffic element and traffic behaviour details. For instance, micro-model describes every single vehicle as the basic unit, when overtaking, car following and changing lanes, the micro-model can be represented.

Macroscopic traffic simulation model has lower description of traffic details, compared with micro-model. Macro-model can describe traffic flow, velocity and traffic density, even the relationship between each other.

Traffic simulation can not only reproduce the traffic situation but also to reflect the real traffic, even more important is that it can be used to test the traffic prediction results or simulate this prediction model.

2.4. Car-following Model

The car-following model is the theoretical basis of the microscopic traffic simulation modeling. In the early 1950s, Reuschel(1950s) and Pipes(1953) [4] began to investigate the car-following model, and the first car-following model was created in 1958 by Chandler.

$$x_n''(t+T) = C\Delta x' \quad (1)$$

x_n'' is acceleration/deceleration for car n, $\Delta x'$ is speed difference with car n and before car, C is sensitivity coefficient, and it is constant number between 0.17-0.74, T is lag time or delay by the reaction time and usually between 1.0s-2.2s.

It is a simple linear model based on stimulus response concept. Subsequently, many car-following model according to this model as a basis to improve and perfect. More well-known is GHR model (by Gazis, Herman and Rothery) [5].

$$x_{n+1}''(t+T) = \alpha \frac{x_{n+1}'(t+T)^m}{(x_n - x_{n+1}(t))^l} * [x_n'(t) - x_{n+1}'(t)] \quad (2)$$

$x_n(t)$ is the displacement of car n at t time.

$x_{n+1}'(t)$ is the velocity of car n+1 at t time, that is following car.

x_{n+1}'' is the acceleration of following car during t time.

m and l are constant number. $m=0\sim 2$ and $l=1\sim 2$.

$\frac{\alpha}{(x_n(t) - x_{n+1}(t))^l}$ is the sensitivity coefficient, and the unit of α is m/s.

Gazis, Herman and Potts proposed that sensitivity coefficient and vehicle's pitch is inversely proportional, also according to the relationship of micro and macro, had evolved this car-following mode.

With the investigation of the car-following model, the researcher awakened to the necessity of driver's stimulation and response in the car-following process. They believe that the transport system cannot simply be seen as a pure mechanical system, but should be seen as an interaction of the system of physical and psychological. Because of the people's subjective factor plays an important role in the system. Therefore, the driver's psychological and behaviour, driving skills, driving purpose, security requirements, apperceiving capability for the traffic and responsiveness to stimuli should be considered in the system. As the diversified development of research tools and data collection, the researchers had integrated cybernetics, statistics, fuzzy math, neural network technology into the car-following model, thus developed the car-following model based on the desired spacing (by Parker and Hidas), driver's psychological response model, fuzzy mathematical models and neural networks.

2.5. Regression Prediction Method

The dynamics of traffic information changes with statistics intervals. When the statistic interval is shorter, so that the traffic time series has a stronger randomness and a weak regularity, the predictability is worse. In contrast, if the statistics interval is longer, the dynamics of traffic information will have more strong regularity and better predictability. Macroscopic traffic information means the transport statistics are counted in a longer term which usually takes a few hours or one day and above. Macroscopic traffic information prediction methods can be divided into quantitative analysis and qualitative analysis methods. The former mainly includes regression analysis prediction methods and time series prediction methods.

The regression analysis prediction method is based primarily on the relationship between the independent variables (factors) and dependent variables (objects) to create a mathematical model. The intending value of dependent variables is forecasted through the independent variables.

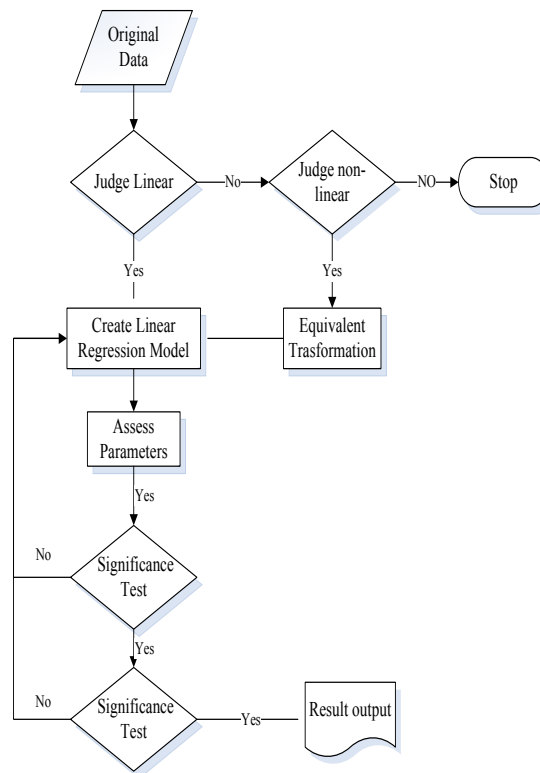


Figure 2.1

The principle of the time series prediction method is forward extrapolation of data's future trends as historical and current data. It mainly includes the deterministic time series and stochastic time series. Overall, the macroscopic traffic information prediction has less historical data and fewer prediction phases. The modelling process needs to consider the impact factor of the macroscopic traffic information.

Regression prediction model can be divided into univariate (unitary) regression model and multivariate regression model or linear regression and non-regression. The regression prediction method process can be summarized as Figure 2.1:

The unitary regression model can be expressed as:

$$y_i = b_0 + b_1 x_i + u_i, \quad i = 1, 2, \dots, n \quad (3)$$

n is the length of measured data.

b is undetermined parameters.

u is random error, it represents a sum of various random factor on dependent variable y .

The independent variable x is influencing factors and dependent variable y is the prediction target.

The multivariate regression model will need to set up a number of influencing factors as x . If the factors have a linear relationship with y , so that the multivariate regression model will be:

$$y_i = b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_jx_{ij} + u_i, \quad i = 1, 2, \dots, n \quad (4)$$

2.6. Traffic Assignment Model

Traffic assignment is an important step in the transportation planning. It means mathematical simulation of the key decisions associated with a distribution channel to compute the optimal solutions regarding transportation [6]. And the start and end points of the trip distribution surveys (OD matrix) assigned to each road in accordance with existing or planned road network, in order to speculate on the road traffic flow.

If the O-D matrix is known and determined, without its time changes, that is call static traffic assignment. It has several prerequisites, which have to know the topological space structure, characteristics, travel time and traffic demand for the road network. The travelers have all the information of the road network and traffic conditions, also need to make the right choice in the same criteria for route choice. In these cases, the traffic assignment will solves the traffic flow of each path in the road network, so that people can evaluate the road network performance.

Dynamic traffic assignment (DTA) needs to know the traffic supply and traffic demand, whereupon the optimal traffic flow distribution pattern will be analyzed. By the control and inducement (guide) the demand of people is reconfigured in space and time, so that the road network can be efficiently run. There is map to show what the relationship between traffic assignment, control and guide.

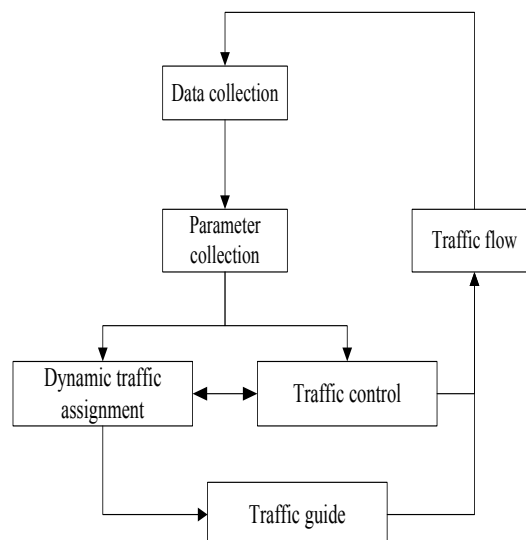


Figure 2.2

Dynamic traffic assignment provides a basis for the traffic control, management and route guidance. The traffic control can reset the signal time of intersection to change the time distribution of traffic flow. The route guidance transfers the traffic information to vehicle guidance system, and using of this non-coercive means to change the spatial distribution of traffic flow. So the most important prerequisite of the dynamic traffic assignment need to

grasp the travel demand around all the time. Then a correct allocation will be generated when traffic demand is absolute known.

3. Future Planning

Traffic simulation and prediction is a very complex research subject, but it can be a huge help for people travel, it is not just conducive to road planning. With further research, we will investigate more models, technologies, theories and methods, and classify and compare the pros and cons between them. Then we hope to create or improve them to adapt the demand of numerous cities or any special city. In this case we want to build a distributed module, and this module will be embedded into vehicle equipment. When the vehicle equipment received the traffic conditions in the front, the module will be analysis the data and forecast the traffic condition in next a few minutes or longer time interval. This process will not be sent to the traffic control centre system, to avoid clogging by the much information generated in control centre. We hope these vehicles' equipment as the input and output terminals. We believe that there is a product acclimatized itself to an era.

Reference

1. <http://baike.baidu.com/>. Baidu Encyclopedia. Traffic control. Traffic condition. Traffic simulation. Traffic assignment model. Traffic prediction. Access Feb/2012. [Website]
2. Traffic control. Traffic simulation. Route assignment. www.wikipedia.org/. Wiki Encyclopedia. Access Feb/2012.
3. Hunlun Xu, Hui Fu. Intelligent prediction theory and methods of traffic information. 2009. Section 1, pp. 40-50.
4. Shoufeng Ma, Guoguang He, Bao Liu. The short-term traffic flow forecasting system Intelligent Transportation Systems, 2004, Vol. 23(2), pp. 28-34.
5. Xiaoyuan Wang, Xinyue Yang, Jinglei Zhang. Microscopic simulation of traffic flow and driver's behavior modeling theory and methods, 2010. pp. 14-28.
6. Chunxia Jin, Hui Wang. Car-following Model and Its Stability Analysis. The institute of systems engineering, ZheJiang University, Hangzhou. 2001, Vol. 1, No. 3.
7. www.entrepreneur.com/encyclopedia. Entrepreneur. Distribution models definition. access Feb/2012.
8. Yu Zhang, Xinghua Hu, Jia Wang. Traffic Planning Model: Application of TransCAD. 2011. pp. 121-142.
9. Haruka S, Prakash R, Takashi N, Yasushi T. An Extended Car-Following Model Combined with Adriver Model. Journal of the Eastern Asia Society for Transportation Studies, Vol. 6, pp. 1545-1556, 2005.
10. TransCAD Transportation demand Model. Chapter 9.
11. Richard W. Rothery. Car Following Models. Civil Engineering Department, The University of Texas, ECJ Building 6.204, Austin, TX78712.
12. Tom Bellemans, Bart De Schutter, Bart De Moor. Models for traffic control.
13. Zhijun Zou, Dongyuan, Yang. Pure-micro car following model for traffic simulation system. Journal of Tongji university 2001, Vol. 29, No. 8.
14. Atul Kumar Sinha, Rahul Maji, Gaurush Hirnandani, Ankesh Kumar Singh. Traffic Control Simulation. Summer Project, 2010, Programming Club, Science and Technology Council, IIT Kanpur.

Special Session

Localisation and Positioning in Sensor Networks

Keynote: Advances in Bayesian methods for Tracking and Decision Making

Dr Lyudmila Mihaylova, Lancaster University

Advances in Bayesian Methods for Tracking and Decision Making

Dr Lyudmila Mihaylova, Lancaster University,
School of Computing and Communications, UK
mila.mihaylova3@lancaster.ac.uk

This talk presents an overview of current achievements, challenges and open avenues of Bayesian sequential Monte Carlo methods in the light of tracking and decision making. The problems are considered first in historical perspective, with the new advances. Special attention is paid to group and extended target tracking. **Groups** are structured objects, formations of entities moving in a coordinated fashion, whose number varies in time. Groups can split, merge, can be near to each other or spawn new groups. Group formations maintain a certain pattern of motion. We distinguish tracking of small groups (with up to 20 components) and large groups (with 100s and 1000s of components). The talk presents an overview of methods, algorithms and techniques especially for high dimensional systems. Recently developed techniques such as Box Particle Filtering are presented, also the core of Finite Random Set Statistics and Markov Chain Monte Carlo methods for high dimensional systems. Figure 1 illustrates the changes in the shape of a Probability Hypothesis Density (PHD) function.

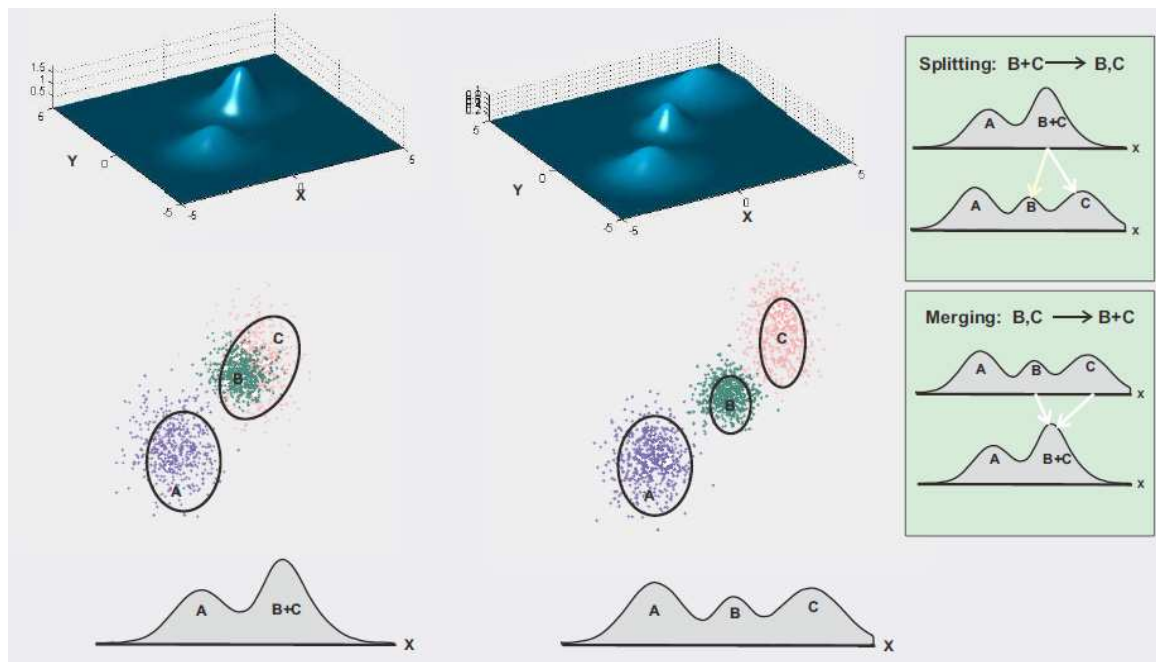


Figure 1: Probability Hypothesis Density function shown as a function of the x and y coordinates. Each peak in the function corresponds to an object.

Examples for localisation and tracking of vehicles in wireless sensor networks are shown on Figures 2 and 3. The base stations are shown and the trajectory of the moving vehicle on google map, from the area of Glasgow, UK. Figure 4 shows results for tracking in sensor networks in the presence of lack of coverage and connectivity. Finally, the challenges related with decision making and behaviour analysis are discussed.

The references below are for: group [1][2][20] and extended object [3][5] tracking, sequential Monte Carlo methods [4][6][7][8][9] with applications to vehicular traffic systems [16], video based tracking [10][11][12][13][14][15] and box particle filtering [18][19].



Figure 2: Localisation in wireless sensor networks

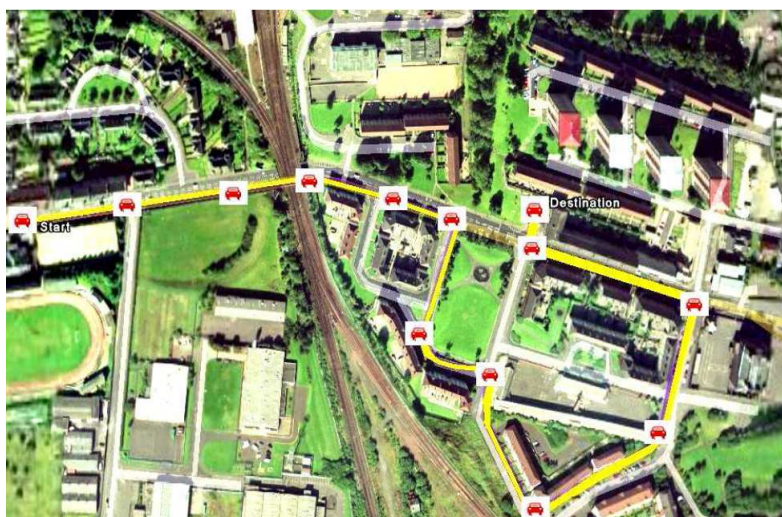


Figure 3: Localisation in wireless sensor networks

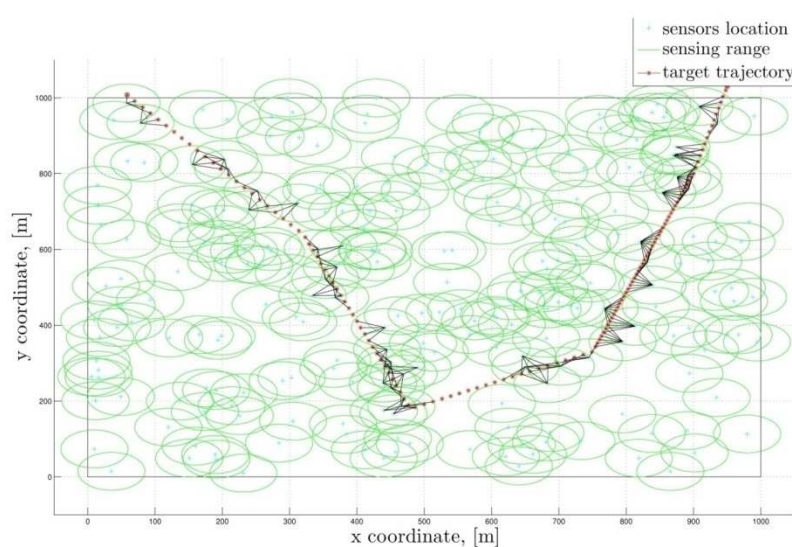


Figure 4: Tracking in wireless sensor networks in the presence of lack of coverage and connectivity

References

1. A. Gning, L. Mihaylova, S. Maskell, S. K. Pang, and S. Godsill, "Group object structure and state estimation with evolving networks and Monte Carlo methods," *IEEE Transactions on Signal Processing*, Vol. 59, No. 4, pp. 1383-1396, April 2011.
2. A. Carmi, L. Mihaylova, A. Gning, P. Gurfil, and S. Godsill, "Markov Chain Monte Carlo based autonomous tracking and causality detection," in *Advanced in Intelligent Signal Processing and Data Mining: Principles and Applications*, P. Georgieva, L. Mihaylova, and L. Jain, Eds. Springer, 2012, pp. 7-49.
3. D. Angelova and L. Mihaylova, "Extended object tracking using Monte Carlo methods," *IEEE Transactions on Signal Processing*, Vol. 56, No. 2, pp. 825-832, 2008.
4. L. Mihaylova, R. Boel, and A. Hegyi, "Freeway traffic estimation within recursive Bayesian framework," *Automatica*, Vol. 43, No. 2, pp. 290-300, February 2007.
5. N. Petrov, L. Mihaylova, A. Gning, and D. Angelova, "A novel sequential Monte Carlo approach for extended object tracking based on border parameterisation," in *Proc. of 14th International Conf. on Information Fusion*, USA, 2011, pp. 306-313.
6. L. Mihaylova, D. Angelova, D. R. Bull, and N. Canagarajah, "Localization of mobile nodes in wireless networks with correlated in time measurement noise," *IEEE Transactions on Mobile Computing*, Vol. 10, pp. 44-53, 2011.
7. L. Mihaylova, D. Angelova, S. Honary, D. Bull, C. Canagarajah, and B. Ristic, "Mobility tracking in cellular networks using particle filtering," *IEEE Transactions on Wireless Communications*, Vol. 6, No. 10, pp. 3589-3599, 2007.
8. L. Mihaylova, D. Angelova, and A. Zvikhachevskaya, "Sequential Monte Carlo methods for localisation in wireless networks," in *Advances in Intelligent Signal Processing and Data Mining: Principles and Applications*, P. Georgieva, L. Mihaylova, and L. Jain, Eds. Springer, 2012, pp. 89-115.
9. L. Mihaylova, A. Hegyi, A. Gning, and R. Boel, "Parallelized particle and Gaussian sum particle filters for large scale traffic systems," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 2012, No. 1, pp. 36-48, 2012.
10. H. Bhaskar and L. Mihaylova, "Combined data association and evolving population particle filter for tracking of multiple articulated targets, article id 642532," *EURASIP Journal on Image and Video Processing*, Vol. 2011.
11. H. Bhaskar, L. Mihaylova, and A. Achim, "Video foreground detection based on symmetric alpha-stable mixture models," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 20, No. 8, pp. 1133-1138, Aug. 2010.
12. A. Loza, L. Mihaylova, D. Bull, and N. Canagarajah, "Structural similarity-based object tracking in multimodality surveillance videos," *Machine Vision and Applications*, Vol. 20, No. 2, pp. 71-83, Feb. 2009.
13. D. Angelova and L. Mihaylova, "Contour segmentation in 2D ultrasound medical images with particle filtering," *Machine Vision and Applications*, Vol. 22, pp. 551-561, 2011.
14. L.A. Klein, L. Mihaylova, N.-E. El-Faouzi, *Sensor and Data Fusion: Taxonomy, Challenges and Applications*, *Handbook on Soft Computing for Video Surveillance*, Editors S.K. Pal, A. Petrosino and L. Maddalena, Chapman & Hall / CRC Cryptography and Network Security, CRC Press, Taylor & Francis Group, 2012, pp. 139-183.
15. A. Loza, L. Mihaylova, F. Wang and J. Yang, *Structural Information Approaches to Object Tracking in Video Sequences*, INTECH book "*Object Tracking*", Edited by H. Goszczynska, I. Lorkovic, T. Smiljanic, M. Sirotic, pp. 23 - 48, Feb. 2011, ISBN: 978-953-307-360-6.
16. A. Gning, L. Mihaylova, R. Boel, *Interval Macroscopic Models for Traffic Networks*, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 12, No. 2, pp. 523-536, 2011.
17. A. Gning, L. Mihaylova, F. Abdallah, and B. Ristic, "Particle filtering combined with interval methods for tracking applications," in *Integrated Tracking, Classification, and Sensor Management: Theory and Applications*, M. Mallick, V. Krishnamurthy, and B.-N. Vo, Eds. John Wiley & Sons, 2012.
18. A. Gning, B. Ristic, L. Mihaylova, A. Fahed, *Introduction to the Box Particle Filtering*, *IEEE Signal Processing Magazine*, 2012, in press.

19. A. Gning, B. Ristic, L. Mihaylova, Bernoulli/ Box-Particle Filters for Detection and Tracking in the Presence of Triple Measurement Uncertainty, *IEEE Transactions on Signal Processing*, Vol. 60, No. 5, pp. 2138-2151, 2012.
20. P. Georgieva, L. Mihaylova, L. Jain (Editors), *Advances in Intelligent Signal Processing and Data Mining: Principles and Applications (Studies in Computational Intelligence)*, Springer-Verlag, 2012, ISSN 1860-949X.

Sensor Position Management based on Interpolation in Wireless Sensor Networks

Arshad Ali¹, Costas Xydeas², Lyudmila Mihaylova³

Department of Computing & Communications

University of Lancaster, Lancaster, UK

¹a.ali1@lancaster.ac.uk, ²c.xydeas2@lancaster.ac.uk, ³mila.mihaylova3@lancaster.ac.uk

Abstract: Wireless Sensor Networks (WSNs) are amongst the most important of the new emerging technologies and have shown an explosive growth in recent years for monitoring physical phenomena. Large scale WSNs are known to suffer from coverage holes, large regions of deployment area where no sensing coverage can be provided. This could happen due to hardware failure, costs for deployment or redeployment. Coverage holes can affect the accurate representation of the natural phenomena that is monitored by WSN. This paper introduces a method to overcome the coverage holes problem by using an interpolation in those areas. It is shown that a phenomenon can be interpolated with high level of accuracy by using the available data coming from different nodes. However, due to energy limitations of sensor nodes it is imperative to perform this interpolation in an energy efficient manner that minimizes communications among nodes. An efficient technique for sensor node positioning is developed based on the following steps. First, we build a correlation model of the phenomena being monitored in a distributed manner. Next, a distributed interpolation technique based on the Kriging interpolation phenomena inside coverage holes is proposed. On the basis of estimated and calculated metric the sensor positions are optimized.

Keywords: Wireless Sensor Network, Kriging, Sensor Positioning, Experimental Variogram.

1. Introduction

Wireless Sensor Networks (WSNs) are very promising for monitoring extraordinary diverse environments. A WSN is made up of wireless nodes that are able to sense some physical information, process and transfer it to its neighbors by establishing wireless ad-hoc network. A WSN consists of tiny sensor nodes each capable of sensing some phenomenon, process data and communicate with its neighbors [1]. These tiny sensor nodes are deployed in the experimental field in large numbers and they collaborate to form ad hoc networks. A node in a sensor network is generally performing two tasks simultaneously, sensing and communicating.

One of the primary challenges for WSNs deployment is the life of the network. As network of this type are normally battery powered, large scale, dispersed over large geographic area and often are out of reach or hazardous environment, maintenance (such as battery replacement) is not feasible.

The life span must be maximized without any significant degradation of the sensory data quality, e.g. by the awake and sleep technique and communicating data in an intelligent way. This paper proposes an approach based on interpolation techniques. Many interpolation models exist that vary in complexity and accuracy. However, we focus on the Kriging technique because it is powerful and we show that Kriging can be used for interpolation in those areas where the sensors are unable to sense the physical phenomena.

The paper is organized as follows. In section 2 we present the related work review on Kriging and sensor networks. In section 3 the Kriging Interpolation technique is explained in details, the problem formulation and solution is described in section 4. Finally, section 5 is devoted to a discussion of the main results obtained in this research work, followed by the appropriate conclusions.

2. Related Work

In the past few years energy efficiency in WSNs has received significant attention. Research work on the configuration of a network topology, with good (or required) connectivity, by using minimal power consumption (such as minimizing the maximum power of nodes or minimizing the total power consumption of all nodes), has been done in [13][14][15]. One of the methods to reduce power consumption and to solve the coverage problem is by means of interpolation techniques. The main attention of this paper is on Kriging interpolation.

Kriging is a statistical tool developed by Matheron [10] and named in honour of D.G. Krige. Originally the Kriging technique was developed for mining and geology. At its simplest form, Kriging can be thought of as a way to interpolate spatial data as an automatic contouring program would. In a more precise manner, Kriging can be defined as an optimal linear unbiased estimator of a spatial variable at a particular site or geographic area. Kriging assigns low weights to distant samples and vice versa, but also takes into account the relative position of the samples to each other and the site or area being estimated. The Kriging technique was very popular in geology and mining as well as to hydrology started in mid-60s and especially in 70s with work done by Georges Matheron. It was the first time when Kriging is applied in such area of science and the estimated values obtained by Kriging are shown to be more accurate than the values estimated by considering indirect indicators (based on the composition of grassland obtained from remote sensing land cover classification data). Kriging is used to interpolate 3-D magnetic resonance imaging (MRI) data in [9]. The authors of [9] observed that Kriging is a preferable interpolation method, because it provides the user with an estimate of the interpolation error introduced. In the field of remote sensing, the detection of changes in land covers the help of remote sensing images in topic of many studies. Cheng et al. [3] used Kriging for image registration. They observed that the Kriging approach gives better results in comparison with polynomial trend mapping. Pham and Wagner [11], proposed a new technique based on Kriging for calculating the linear prediction coefficients for speech. They observed that this new technique is efficient due to the unbiased behaviour of Kriging and it could provide means for adaptive linear prediction coding. Due to the use of theoretical semivariogram, it is shown to provide good interpolation values of the number of poles for linear prediction coding [6]. In [8], Kriging interpolation is used in WSNs to find the optimal number of sensors in cool chain management to monitor the temperature in delivery truck. The method is used to estimate the minimum number of sensors and to compare different sensor positioning strategies [8].

3. Kriging Interpolation Techniques in WSNs

Often, it is not feasible to deploy a WSN with high density in order to be able to observe desired phenomena at every point in the sensing field. Factors that influence this feasibility are the deployment cost, the physical size of the device and the life of the sensor node. Therefore, to overcome these problems a mechanism needs to be introduced in order to interpolate data at intermediate points between sensor nodes.

A number of recent works in the WSN data acquisition domain propose to exploit this spatial correlation to perform energy efficient data collection [12]. In traditional monitoring domains, such as remote sensing, estimation methods based on spatial correlation are commonly used to interpolate various spatial phenomena [5]. In this work we use the spatial correlation of WSN data to interpolate temperature values at locations that are not covered by sensors (see Figure 1 for an example). Simple approximation methods are available as well, e.g. computation of the sample average could be applied as spatial interpolation. However, such methods are often quite inaccurate. The spatial averaging method includes simple and inverse distance weighted averages or assumes uniform spatial correlation [12]. In this study, the Kriging technique [14] is of interest.

Kriging methods are a family of techniques that can interpolate a random field at unobserved locations by using values of neighbour locations. Kriging is based on the assumption that the parameter being interpolated can be treated as a regionalized variable. A regionalized variable is intermediate between a truly random variable and a completely deterministic variable. It varies in a continuous manner from one location to the next and therefore points that are near each other have a certain degree of spatial correlation, but points that are widely separated are statistically independent. Kriging is a set of linear regression routines which minimize estimation variance from a predefined covariance model.

$$\hat{Z}(x_0) = \sum_{i=1}^n (\lambda_i Z(x_i)) \quad (1)$$

where Z is the measured event at point x_i , λ_i is the kriging weight, \hat{Z} is the interpolated data at point x_0 .

There are several Kriging techniques for different purposes such as ordinary Kriging, simple Kriging, block Kriging, Bayesian Kriging, etc. In this research we use the ordinary Kriging as it is the most common type of Kriging. The most common Kriging type of methods are:

1. Simple Kriging, when the mean $[m(Z_0)]$ is assumed to be known and constant over the whole study area [2].
2. Ordinary Kriging, if the mean $[m(Z_0)]$ is unknown but considered as constant in a local neighbourhood [2].
3. Kriging with a trend model (or Universal Kriging), when one assumes that the unknown mean $m(Z_0)$ is not constant. It can then be modelled as a linear combination of functions of the coordinates [2].

3.1 Variogram Modelling

This section presents the principles of variogram modelling and relevance to the Kriging technique. The problem of fitting the best variogram model is presented. In order to use the Kriging technique the covariance model of matrix of the random variable is needed [4]. The variogram is defined as the expected squared difference between two data points separated by a distance h . A typical measure of statistical distance in spatial phenomena is the experimental variogram (EV, $\gamma(h)$) defined as a function of samples of a phenomenon and distance between the corresponding sampling locations [7].

Assume a random variable Z represents a spatial phenomenon and $Z(x)$ represents a sample at location x then for a given distance h (lag), the EV is defined as:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{N(h)}^+ [Z(x) - Z(x+h)]^2 \quad (2)$$

where, $N(h)$ is the number of data pairs at distance h . The spatial correlation in a phenomenon that can be observed by computing the EV values for several lags using available samples. However, EV computation alone is not sufficient for Kriging as this method often requires correlation between locations where no samples are available. For this purpose, a spatial correlation model is established by fitting a curve onto the computed EV values. The resulting spatial correlation or variogram model is used to find the correlation between phenomenon values at arbitrary locations.

3.2. Choosing an Appropriate Variogram Model

The variogram model for a phenomenon can be established by choosing a model that best fits its EV values. However, in practice this choice is limited by an important feature of the Kriging system of linear equations SLE [7] i.e. a unique solution of the Kriging SLE is only possible if the Kriging matrix (matrix A in $A^{-1}b$) is non-singular and positive definite [7]. It is not possible to use an arbitrary curve fitted onto the EV values as a variogram model as such a model cannot guarantee positive certainty of the resulting Kriging matrix. Hence, in

practice only some well-known functions are used as variogram models [7]. A typical choice of variogram model is the Gaussian distribution.

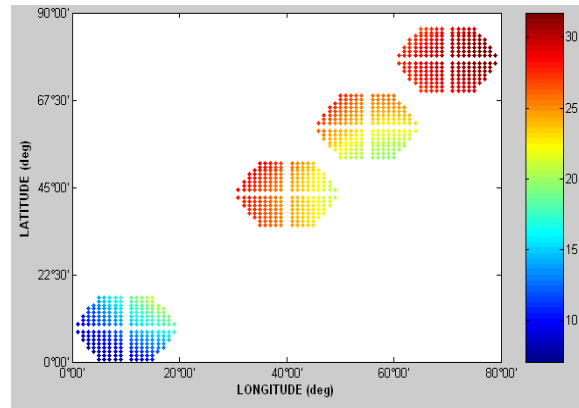


Figure 1: Temperature measurement with 4 sensors

The Gaussian Model

If the variance is very smooth and nugget variance is very small compared to the spatially dependent random variation, then the variogram can often be best fitted with a Gaussian model

$$\gamma(h) = n + s \times \left(1 - \exp\left(-3 \frac{h^2}{a^2}\right) \right) \quad (3)$$

The result is shown on Figure 2.

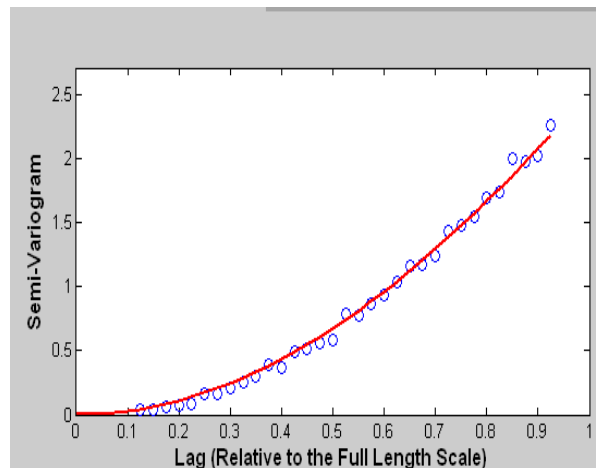


Figure 2: Variogram fitting

The Kriging Interpolation Method

Once a variogram model is established for a given phenomenon, it can then be used for spatial interpolation using Kriging. In ordinary Kriging, a spatial phenomenon Z is assumed to be represented by its realization $Z(x_1), Z(x_2), \dots, Z(x_n)$ at locations x_1, x_2, \dots, x_n , then the Kriging interpolator of Z at a point x_0 is given by [6]

$$\hat{Z}(x_0) = \sum_{i=1}^n (\lambda_i Z(x_i)), \quad (4)$$

where λ_i are the weights fulfilling the unbiasedness condition, i.e., $\sum_{i=1}^n (\lambda_i) = 1$ and the expected error is $E [\hat{Z}(x_0) - Z(x_0)] = 0$ [7]. Kriging is an optimal estimator in the sense that it minimizes the estimation variance and is unbiased [8]. It can be shown that optimal weights λ_i for the Kriging interpolator can be computed from the following system of linear equations [7].

$$\Lambda = A^{-1}B$$

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} \gamma(x_1, x_1) & \cdots & \gamma(x_1, x_n) \\ \vdots & & \vdots \\ \gamma(x_n, x_1) & \cdots & \gamma(x_n, x_n) \end{pmatrix}^{-1} \begin{pmatrix} \gamma(x_1, x^*) \\ \vdots \\ \gamma(x_n, x^*) \end{pmatrix} \quad (5)$$

where Λ is a vector comprising of Kriging weights λ_i and a Lagrange multiplier (added for computational reasons), 'A' is the spatial correlation matrix of sample locations x_1, x_2, \dots, x_n and b is a vector whose elements represent the spatial correlation between x_0 and each $x_i \{x_1, x_2, \dots, x_n\}$. All correlations are based on an appropriate variogram model defined for the spatial phenomenon under observation [7].

4. Problem Formulation

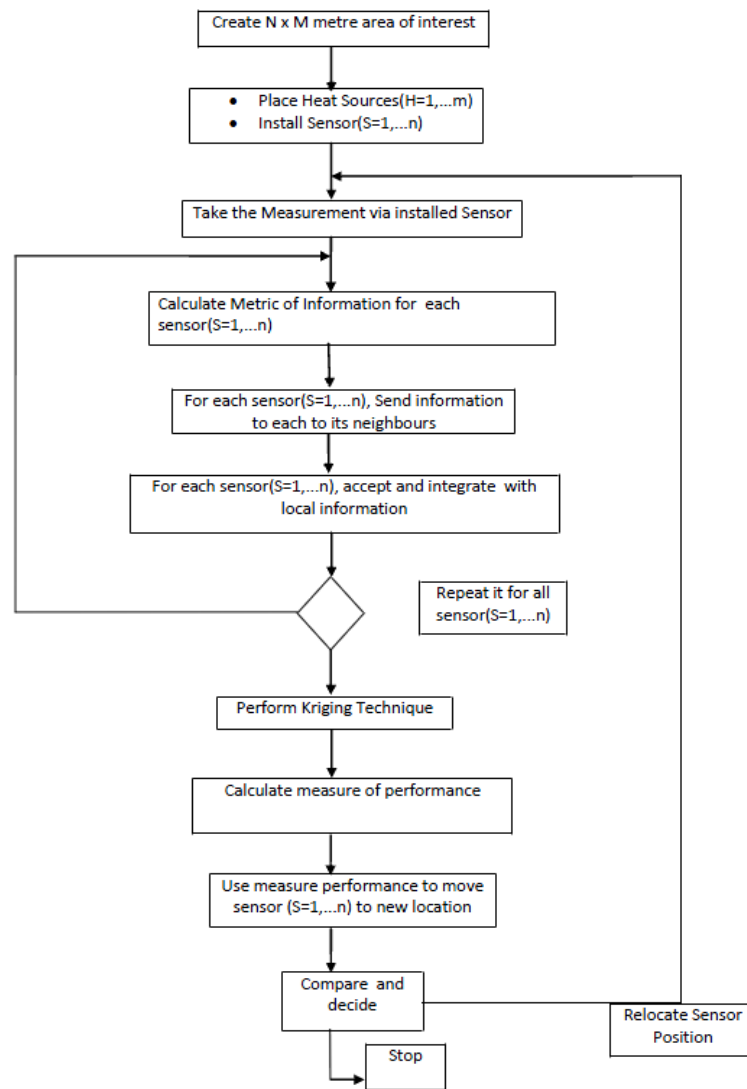
One popular application of WSNs is to monitor a given physical phenomena in an area of interest generally called a sensor field. After having established the network, each node is capable to monitor a region depending of its sensing range. The sensing and communication range of a sensor node depends on its corresponding power and nature. Ideally, the sensor network is supposed to be able to monitor the entire sensing field by combining all the pieces of information coming from its sensor nodes. However, for large region, this total coverage demands a large number of sensor and therefore a costly and energy demanding system. To overcome this problem, a limited and reasonable number of moving sensors is a natural solution.

In this study, for a given number of moving sensors, we propose a method to find the optimal positioning of sensor nodes in order to extract maximum information about the environment. A metric based on gradient method is used to guide the sensors displacements toward the optimal placement in the sense of estimating the desired physical phenomena. At each time step, the currently available measurements along with the previous time measurements are used to predict metric values in the entire region of interest. This metric prediction is performed using a Kriging technique until a condition of convergence is reached.

4.1. Definition of the Environment and the Sensor Deployment

Ideally, the sensor networks should be able to monitor the entire sensing field by combining all the pieces of information coming from its sensor nodes. However, for large regions, this total coverage demands a large number of sensor and therefore a costly and energy demanding system. To overcome this problem, a limited and reasonable number of moving sensors is a natural solution. The area of interest is (N x M) meters and the physical phenomena to be monitored is the temperature. In the current study, the temperature is supposed to be static in relation to the time evolution for simplicity.

System Block Diagram



To simulate the spatial distribution of the temperature, ‘n’ number of heating sources are placed in the sensor field (see Figure 4). Each heat source $(H_i)_{i=1,...,n}$ is assumed to create a Gaussian distributed temperature centred on its position A_i , with a mean T_i and standard deviation σ_i . Following this Gaussian distribution chosen for the heat sources, the temperature at any point $M = (x, y)$ if we assume the presence of only one heat source is given by

$$T_{A_i} = T_i \left(\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left(-\frac{|A_i - M|^2}{2\sigma_i^2} \right) \right) \quad (6)$$

Then, the temperature T at any point M with n numbers of heat sources is given below:

$$T = \sum_{i=1}^n T_{A_i} \quad (7)$$

Furthermore, m moving sensor nodes are initially randomly deployed in the sensor field. It is assumed that each sensor takes uniformly several measurements inside a hexagon with the

half diagonal taken to be the sensing range r_i . The hexagonal shape is chosen since it allows covering entirely the squared region of interest (see the Figure 3).

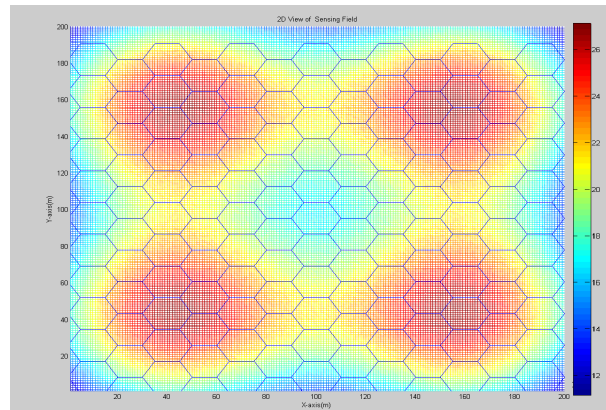


Figure 3: 2D view of the spatial distribution of the temperature with finite number of hexagons.

As prior information, the squared region can be subdivided into a finite number of hexagons. The hexagon centres define possible positions of the deployed sensors. In addition, in one step, the sensor movements are constrained to the six neighbour hexagons around the current position. As stated previously, inside each hexagon, each sensor is assumed to uniformly take several measurements used for the calculation of a metric that is defined in the next paragraph 0.

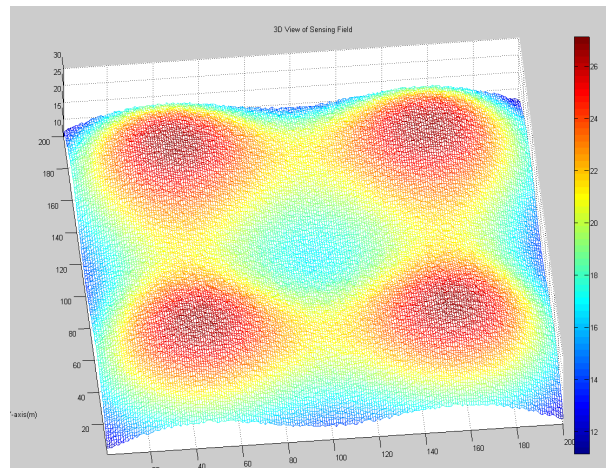


Figure 4: 3D view of the spatial distribution of the temperature

4.2. Metric of Information

The gradient of a specified function is used as a measure of performance for sensor position optimization in a global framework in the field of interest. Each sensor calculates a gradient value by using a uniform grid inside its current hexagon position. Next, each sensor broadcasts its local metric to its neighbors along with its current position coordinates. Note that, in the current study, for simplicity, all sensors are direct neighbors. All the sensors integrate all the new information in its memory table which also contain all the information from the past. Finally, the total information is used for a Kriging interpolation of the metric in the region of interest.

After performing this interpolation, the next step consists on deciding the next movement of the sensors. This step is described in the next section.

4.3. The Proposed Algorithm for Optimal Sensor Positioning

In the first iteration each sensor has m (the number of sensors) metrics to perform the Kriging estimation. On the k^{th} iteration each sensor memory table contains $k*m$ metrics used to perform Kriging. Therefore during the time evolution, each sensor gathers more and more information about its environment that is used for enhancing both the global estimation and the sensor placement in the sensor field. The developed algorithm is described with the pseudo code given below.

```

t:= 0; Initialize (Sen_Node, H_Sour);
for:
  take measurement;
  process data;
  find metric of performance;
  send to neighbors;
  for:
    perform kriging to enhance vision;
  find out area of interest closer to each sensor
    relocate one step towards interested area

  end; compare  $\frac{\sum_{i=1}^n M}{n} < \xi$ 
end; stop;
    
```

5. Current Investigation

The ' m ' number of sensors are installed in the area of interest and each sensor is supposed to measure temperature from the experimental field. On the basis of the measured temperature each sensor calculates metric of information by using measured data and metric based environment is shown in Figure 5.

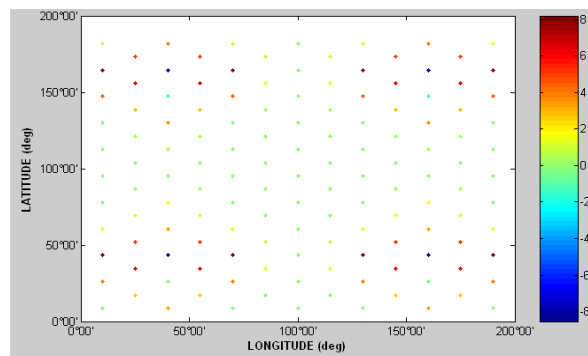


Figure 5: Two dimensional metric based environment

The Figure 6 below is showing the measured and estimated metric difference by using the Kriging interpolation technique and it is showing that by increasing the number of measured metric the error between measured and interpolated metric is reduced.

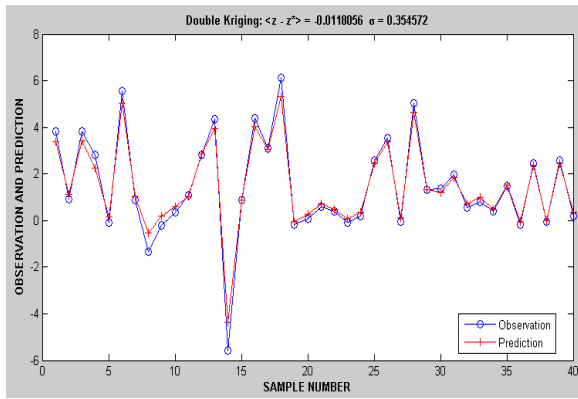


Figure 6: Measured versus estimated Metric

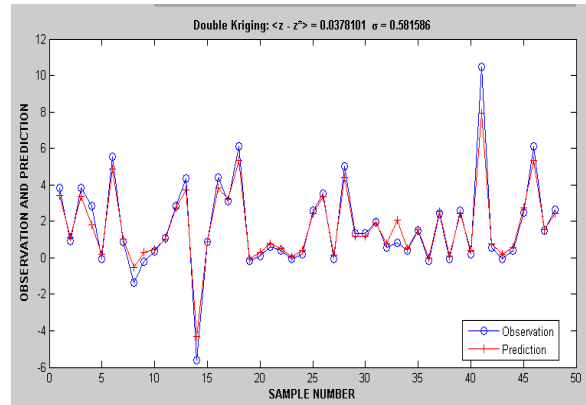


Figure 7: Measured versus estimated metric

The Figure 8 below is showing that the error difference between measured and interpolated metric on each iteration, in 1st iteration with ' m ' number of measured metrics and in 2nd iteration ' $2m$ ' number of measured metrics and so on. In the k^{th} iteration we have iteration ' $k*m$ ' number of measured metrics and it has comparably very less error with respect to previous iterations.

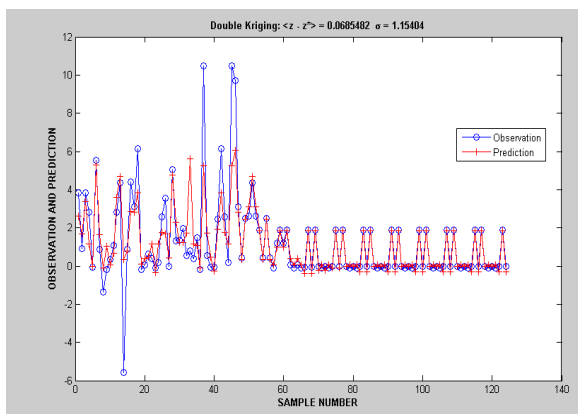


Figure 8: Measured versus estimated metric

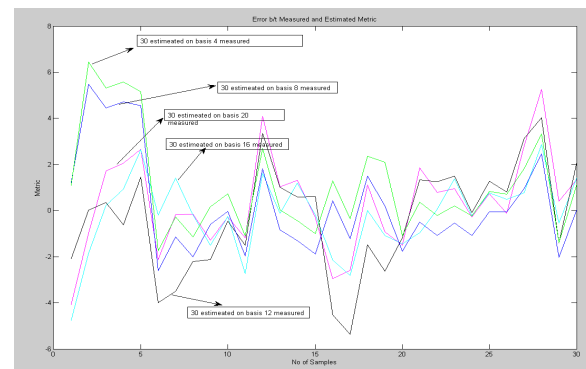


Figure 9: Error between measured and estimated metric

In the Figure 10, average error is plotted and it is average error between ' m ' and ' $m-1$ ' number of iteration and it is showing that it is decreasing while number of iteration is increasing (number of measurements).

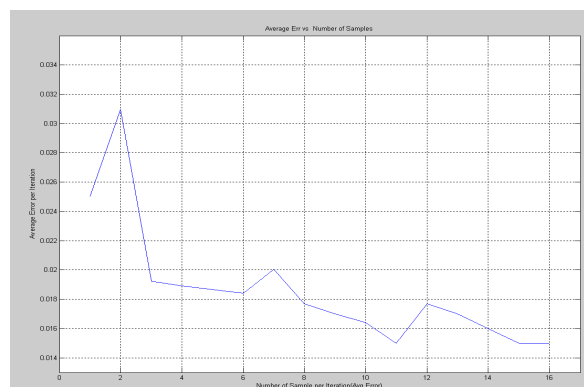


Figure 10: Average error

In current work is the more focus is on Kriging Interpolation and according to the above results, it is showing that it is converging while collecting more and more data from the sensing field. At the time being sensor are moving to the maximum and minimum metric for new location.

6. Conclusions

As coverage holes are the real problem in WSNs, it is important to interpolate phenomena in a coverage hole. To address this problem, spatial interpolation is best spatial correlation model is common in WSN. The challenge is to perform the interpolation with high accuracy and minimum communication cost. To overcome this problem we used to two step technique, we first models the spatial correlation in WSN by using the processed data and after that interpolates data in the coverage holes. Our interpolation technique, distributed and localized Kriging method is highly scalable as compared to centralized interpolation technique for node positing in the field.

The future work will be focused on techniques using metric based map for sensor positing, optimal global view of the sensing field and more complex scenarios.

References

1. N. Ahmed, S. Kanhere, S. Jha, The Holes Problem in Wireless Sensor Networks, A Survey, pp.4 – 18, *ACM SIGMOBILE Mobile Computing and Communications Review*, Vol. 9, No. 2, 2005.
2. R. P. Barry, R. K. Pace. Kriging with Large Data Sets Using Sparse Matrix Techniques, *Communications in Statistics - Simulation and Computer*, Vol. 26, No. 2, pp. 619-629, 1997.
3. K. S. Cheng, C. Wei, S. C. Chang, Locating Landslides Using Multi-temporal Satellite Images, *Pattern Recognition Society*, Vol. 41, No. 11, pp.1744-1754, 2008
4. D. Chu, A. Deshpande, J. Hellerstein, W. Hong approximate data collection in sensor networks using probabilistic models, *IEEE Intl. Conf. on Data Engineering*, pp. 48-60, 2006.
5. P. J. Curran, PM Atkinson, *Geostatistics and Remote Sensing Progress in Physical Geography*, Vol. 22, No. 1, pp. 61-78, 1998.
6. X. Emery, Simple and Ordinary Kriging Multigaussian Kriging for Estimating recoverable Reserves, *Mathematical Geology*, pp. 295-319, 2005.
7. E. H. Isaaks, R. M. Srivastava, G. H Schubert, *An Introduction to Applied Geostatistics*. Oxford University Press, New York, USA. Canadian Journal of Forest Research, pp.1399-1407, 1999.
8. R. Jedermann, W. Lang, The Minimum Number of Sensors – Interpolation of Spatial Temperature Profiles in Chilled Transports, *LNCS*, Vol. 5432, pp. 232-246, 2009.
9. S. M. Matechik and M. R. Stytz. Using kriging to interpolate MRI data, *New Opportunities for Biomedical Engineers, Proc. of the 16th Annual International*, 1994.
10. G. Matheron, *Principles of Geostatistics*, Economic Geology, Vol. 58, pp. 1246-66, 1963.
11. T. D. Pham and M. Wagner, A Geostatistical Model for Linear Prediction Analysis of Speech, *Pattern Recognition Society*, Vol. 31, No. 12, pp. 1981-1991, 1998.
12. A. Silberstein, R. Braynard, J. Yang, Constraint chaining: on energy-efficient continuous monitoring in sensor networks, *Proc. of SIGMOD*, pp. 157-168, 2006.
13. R. Ramanathan, R. Rosales-Hain, Topology Control of Multihop Wireless Networks Using Transmit Power Adjustment, *Proc. IEEE*, Vol. 2, pp. 404-413, 2002.
14. S. Singh C.S. Raghavendra and J. Stepanek, Power-Aware Broadcasting in Mobile Ad Hoc Networks, *Proc. IEEE 10th International Symp.*, 1999.
15. R. Wattenhofer, L. Li, P. Bahl, and Y.M. Wang, Distributed Topology Control for Power Efficient Operation in Multihop Wireless Ad Hoc Networks, *Proc. of the IEEE International Conf. Comm.*, 2001.

Stochastic Model for Improving Connectivity Under Heterogeneous Traffic Flow for VANET

¹Tariq.Umer, ²Hassan Ahmad, ³Bahram Honary

School of Computing and Communications, Lancaster University,
Lancaster, United Kingdom

¹t.umer@lancaster.ac.uk, ²h.ahmad@lancaster.ac.uk, ³b.honary@lancaster.ac.uk

Abstract: The traffic flow conditions and the vehicular characteristics affect the connectivity in Vehicular Ad-hoc Network (VANET). In this paper, a fluid dynamic model has been applied to incorporate heterogeneous traffic flow having the influence of microscopic parameters of traffic flow theory. The model exhibits the changes in velocity and density of vehicles under the effect of jamming density, while keeping vehicles held apart at a fixed safety distance from each other required by traffic authority laws. In order to maintain connectivity in VANET, a dual ring connectivity scheme has been introduced using the Warning Energy Aware Cluster-head (WEAC) protocol. Dual ring of network communication is established, forming a primary ring by using small vehicles whereas creating a secondary ring of long vehicles as backup path for multi-hop connectivity. The model have cluster-head based routing algorithm, which gives improved connectivity dynamics of the network when having backup path for communication.

Keywords: Connectivity, Traffic flow model, Density estimation, Vehicular Ad-hoc network.

1. Introduction

Traffic flow theory has introduced traffic codes for the safety and smooth flow of traffic on the roads [1]. Traffic engineers have defined different traffic flow models based on both microscopic and macroscopic parameters to implement traffic codes. These models define relationship between the characteristics of moving vehicles such as speed, density flow, safety distance, and length [2]. Researchers have made use of these underlined parameters, to analyse the road conditions and vehicular characteristics for their area of interest.

Design and implementation of the vehicular ad-hoc network depends on the prevailing road conditions. The vehicular characteristics of moving traffic under diverse road conditions are used by research studies to focus the key issues of VANET. The use of density dynamics for the assignment of dynamic transmission range in vehicular ad-hoc network was performed in these studies [3][4]. The mobility of vehicles on the road is also focused to enhance the performance of VANET. The available models of traffic flow theory such as the fluid dynamic model and the car following model are used for representing traffic flow in many studies for VANET [5]. In [6], analytical modelling of VANET by considering the traffic flow models and analysing the statistical properties of connectivity is well elaborated.

The use of queuing theory and consideration of road traffic parameters for connectivity analysis are discussed and presented as analytical model for connectivity in [7]. In [8][9] the main focus is on study of connectivity under vehicles speed, distribution, cluster formation, distance between the vehicles and power consumption. The authors in [10] have made use of percolation theory results and presented simulation model while considering traffic lights, road side infrastructure, transmission range and proportion of equipped vehicles. The study [11] has presented a stochastic traffic model for road traffic that focuses on connectivity in signalized road systems by using the fluid model for density estimation and then using stochastic model for analysing the connectivity issue.

In [12] to observe vehicles density dynamic under more realistic road environment, we have implemented new heterogeneous traffic flow environment under the effect of safety distance between the vehicles. The model previously presented in [13] for density estimation captures

less realistic road conditions for density estimation. The proposed model reports the density of moving vehicles on the highway as calculated from the fluid dynamic model having the effect of road conditions under heterogeneous traffic flow. From the fluid dynamic model defined in partial differential equation (PDE), we have mean number of vehicles and mean flow rate. To get the actual number of vehicles within a certain travelled area we have used Poisson Arrival.

Location Model (PALM) presented in [14][15]. We have implemented the stochastic model for the heterogeneous traffic flow under the influence of microscopic parameters.

In our connectivity model, we have made use of key feature of heterogeneous traffic flow that is, diversity in types of vehicles available on the road. The concept of dual-ring for communication is introduced. On the highway or two lane urban road environments, the fast lane vehicles will be forming a primary ring of communication, whereas long vehicles having slow speed moving in slow lane under traffic laws, will be providing secondary ring as backbone for communication. For the implementation of dual-ring concept, we have used power-aware hybrid routing protocol, known as the

Warning Energy Aware Cluster-head (WEAC) protocol [16]. The protocol defined Cluster-head from the moving vehicles for the communication.

The remainder of this paper is structured as follow: In section-2, by using the fluid dynamic model and stochastic model, density estimation is performed. Section-3 defined the connectivity model, the WEAC protocol and the connectivity analysis. Section-4 presents simulation results and their analysis for VANET connectivity. Section-5 provides conclusion of the paper.

2. Density Estimation for Heterogeneous Traffic Flow for Highway

2.1. Using Deterministic Fluid Model

The traffic flow theory defines traffic flow on the road as fluid. The traffic models capture the influence of the fluid like characteristics of traffic to gain the desired outcome for the safe and secure transportation.

The fluid dynamics model resides under the umbrella of continuum traffic flow model and represents the traffic flow in the form of conservation law. It deals with the traffic flow rate and density as a function of time and space and defines these variables in the form of partial differential equation.

In this paper a heterogeneous traffic flow in a Dual-lane, one way and semi-infinite highway twp road environment has been considered. The location space is characterized with the interval $[0, \infty]$, the starting of the road is marked by boundary point 0, which is considered as spatial origin. The road is divided into number of segments represented by $r = 1, 2, 3, 4$.

For the first segment of the road, the number of arrivals of all type of vehicles up to time t , is counted by an arrival process $\{G(t) | -\infty < t < \infty\}$, which is assumed to be finite with probability of 1. This arrival process is characterized by external arrival rate function ' $\lambda_i(t)$ ' for all types of vehicles. The rate function is non-negative and can be integrated. As the two types of vehicles being considered, the arrival rate for cars and busses is given by the external arrival rate function $\lambda_c(t)$ and $\lambda_b(t)$.

The key parameters of traffic flow theory such as vehicles flow and density are well related through the conservation equation given as

$$E^+(x, t) = N(x, t) + F(x, t) + E^-(x, t) \quad (1)$$

Defining $(N(x, t), F(x, t))$ as Total number of vehicles in location $(0, x)$ and the Number of vehicles passing past position 'x', whereas ' $E^+(x, t)$ ' and ' $E^-(x, t)$ ' are respectively, for the vehicles arriving and departing rate. Implementing this conservation equation for cars and buses

$$E_c^+(x, t) = N_c(x, t) + F_c(x, t) + E_c^-(x, t) \quad (2a)$$

$$E_b^+(x, t) = N_b(x, t) + F_b(x, t) + E_b^-(x, t) \quad (2b)$$

By applying the operator $\frac{\partial^2}{\partial x \partial t}$ on the equation (2a, 2b) which differential the equation (2a, 2b) by time and space, we will have partial differential equation form of conservation equation [13].

$$\frac{\partial n_c(x, t)}{\partial t} + \frac{\partial f_c(x, t)}{\partial x} = e_c^+(x, t) - e_c^-(x, t) \quad (3a)$$

$$\frac{\partial n_b(x, t)}{\partial t} + \frac{\partial f_b(x, t)}{\partial x} = e_b^+(x, t) - e_b^-(x, t) \quad (3b)$$

According to the fundamentals relation of traffic flow theory

$$f(x, t) = n(x, t) u(x, t) \quad (4)$$

Using equation (2(a, b)) and (3(a, b)) we have

$$\frac{\partial n_c(x, t)}{\partial t} + \frac{\partial [n_c(x, t) u_c(x, t)]}{\partial x} = e_c^+(x, t) - e_c^-(x, t) \quad (5a)$$

$$\frac{\partial n_b(x, t)}{\partial t} + \frac{\partial [n_b(x, t) u_b(x, t)]}{\partial x} = e_b^+(x, t) - e_b^-(x, t) \quad (5b)$$

These relations formed the one dimensional version of generalized conservation law for fluid motion in partial differential form representing cars and busses in heterogeneous traffic flow.

By applying chain rule and defining velocity as

$$u(x(t), t) = \frac{dx(t)}{dt}$$

After substituting values from (4) we get equation for finding density of two different type of vehicle.

$$\frac{dn_c(x(t), t)}{dt} = e_c^+(x, t) - e_c^-(x, t) - \frac{\partial u_c(x, t)}{\partial x} \cdot N_T(x(t), t) \quad (6a)$$

$$\frac{dn_b(x(t), t)}{dt} = e_b^+(x, t) - e_b^-(x, t) - \frac{\partial u_b(x, t)}{\partial x} \cdot N_T(x(t), t) \quad (6b)$$

$$N_T(x, t) = n_c(x(t), t) + n_b(x(t), t) \quad (7)$$

This equation for total vehicle density has effect of microscopic variables of headway and safety distance for different type of vehicles in the traffic stream [17][18].

The velocity profile used in equation 6(a) & 6(b) is defined by the equation (8) presented in our previous paper [12].

$$U_i(x, t) = u_f \left[(1 - N(x, t)) \sum_{i=1}^N p_i (L_i + h_i) \right] \quad (8)$$

2.2. Stochastic Model for Heterogeneous Traffic Flow

The density estimation from fluid dynamics model provides expected number of vehicles. To capture actual number of vehicles within the maximum travelled area, we need additional distribution information from the stochastic model. In stochastic model the density and flow rate for car and bus will be define as partial derivate of expected values.

$$n_c(x, t) = \frac{\partial E[\partial N_c(x, t)]}{\partial x} \quad \& \quad n_b(x, t) = \frac{\partial E[\partial N_b(x, t)]}{\partial x}$$

$$f_c(x, t) = \frac{\partial E[\partial F_c(x, t)]}{\partial x} \quad \& \quad f_b(x, t) = \frac{\partial E[\partial F_b(x, t)]}{\partial x}$$

Under the implementation of PLAM, the arrival process $\{G_c(t), G_b(t) \mid -\infty < t < \infty\}$ for car and bus for the first segment of road is a non-homogenous Poisson process with non-negative and integratable external arrival rate function $\lambda_c(t)$ for car and $\lambda_b(t)$ for bus respectively. So the number of arrivals in the interval (t_1, t_2) is passion with mean

$$\int_{t_1}^{t_2} \lambda_c(w) d(w) \quad \& \quad \int_{t_1}^{t_2} \lambda_b(w) d(w)$$

According to [15][16] we can have $N_T(x, t)$, the total random number of all type of vehicles within the range $(0, x)$ at time 't', via stochastic integration starting with Poisson process (G_c, G_b) where $(G_c(t), G_b(t))$ counts the number of arriving cars and buses to the road segment up to the time t.

$$N_T = \int_{\beta(x, t)} 1_{\{Lw(t) \in (0, x)\}} dG_T(w) \sum_{n=G_T(\beta(x, t))}^{G(t)} 1_{\{Lw(t) \in (0, x)\}} \quad (9)$$

Hence for all real t, $\{N_T(x, t) \mid x \geq 0\}$ is a Poisson process with

$$E[N_T(x, t)] = \int_{\beta(x, t)}^t \alpha(w) d(w) \quad (10)$$

Where G_r is the rth jump time of G , counting backward from time t. 1_c is an indicator function which that turns 1 if C is true and 0 otherwise. $L_w(t)$ is the location process, which defines the position of vehicle on the road segment at time t arrived at time W. Let $\beta(x, t)$ denote the road entrance time for any type of vehicle to be at position x by the time t. for the other vehicles arriving after $\beta(x, t)$ they will be still in position x at time t.

3. Connectivity Analysis

To analyse and achieve a firm connectivity for VANET under more realistic road condition, we have vehicles density estimation from fluid dynamic model having effect of microscopic parameters. The available vehicles in the observed area communicating through a radio link having a transmission range 'R'. The changing road conditions under the influence of heterogeneousness of available vehicles types and traffic laws affect the vehicles density and creating dynamic network topology for vehicle communication. The two vehicles are considered to be connected under these conditions.

Condition 1: A two vehicles are said to be connected if Euclidean distance between them is equal or less then the transmission range i.e. ($D_i \leq R$) where D_i is Euclidean distance between two consecutive nodes.

Condition 2: Vehicles moving in one-dimensional wireless network does not exhibit connectivity from the forward network if any of vehicles is not connected to any forward neighbours.

Condition 3: A wireless network is considered connected if the degree of every node is not equal to 0, if a node has zero minimum node degree ' $D = 0$ ' it will be called isolated and not connected to any neighbour.

3.1. Dual Ring Connectivity Model

The heterogeneousness of vehicles is available for their type, observed traffic laws and moving pattern. These characteristics of moving vehicles provide interesting aspects to deal with the VANET issues. On the road, we can divide vehicles as two available types. The long vehicles defining slow speed, buses, trucks etc and small fast moving vehicles like cars. The road infrastructure and traffic laws bound the two types to move in specific pattern. The fast moving vehicles follow the middle lane and the slow moving vehicles are bound to use side lane.

In our proposed dual ring model, we have considered this specific vehicles road moving pattern to secure connectivity. The vehicles moving on the inner fast lane such as cars are making primary ring of communication while the vehicles available on slow speed lane knows as long vehicles are forming secondary ring. If the vehicles available on travelled area are within the transmission range ' R ' they will be connected through the primary ring. When the distance ' D ' available between two fast moving vehicles is increased such as ' $(D > R)$ ' the vehicles are disconnected. In this scenario having long vehicles on the slow lane the communication path is shifted to secondary ring performing as a backbone for communication network.

To establish the dual ring communication, we have implemented Warning Energy Aware Cluster-head (WEAC) hybrid protocol. This protocol creates a dynamic wireless mobile infrastructure and provides high QoS for highly mobile and heavy loaded Ad-hoc networks. During the communication a mobile node is elected from the set of nominees to act as a temporary base station for a period of time within its zone. In Figure 1 the two different road scenarios of dual-ring implementation are described.

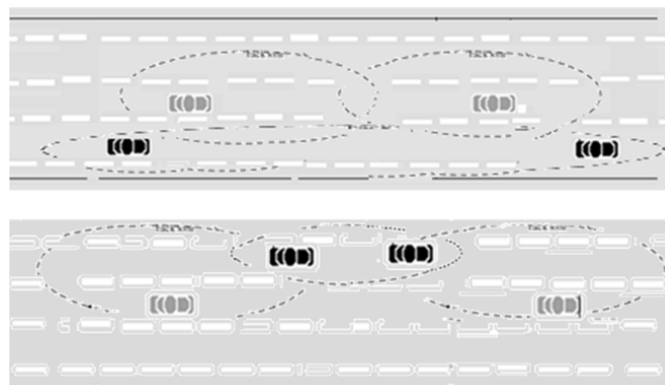


Figure 1: Implementation of Dual Ring in different Road scenarios

3.2. Implementation of WEAC Routing Protocol

As WEAC (Warning Energy Aware Cluster-head) is cluster-based protocol and for the multihop communication in VANET having different type of vehicles on the road, a small vehicle such as car is selected on the primary ring as cluster head whereas for the secondary ring communication long vehicle is selected as cluster head. The available Zone-MT within the operating area of one cluster-head communicates with their neighbours and cluster head through hello packets for updating about its location and neighbours list. Having a power-aware characteristic of a protocol, a node can select its cluster head either on primary or secondary ring. During communication a Zone-MT works under Zone cluster-head as long as it lies in its transmission range.

When the transmitting MT fails to reach its destination through its own Zone cluster-head on primary ring it forward traffic by using available neighbouring Zone cluster-head on secondary ring.

4. Simulation Setup and Results

To analyse the connectivity dynamics with dual-ring infrastructure for heterogeneous traffic flow, we have run the simulation in Matlab under different arriving ratio of vehicles and communication range. In each scenario vehicle looking to transmit on the network established its neighbours list within the define transmission range. In case of vehicle not directly available on the primary ring neighbour list, it is searched through the neighbour list of secondary ring and approached in multi-hop manner. The max-node degree of every case is calculated which defines the projected area of communication under certain transmission range and vehicles density. In Figure 2 the arriving ration for car P_c & bus P_b is fixed as 12/20 and 08/20 for having same heterogeneous traffic environment for all transmission range changing from 50m to 400m. A Dual lane highway environment is implemented, which provides two-dimension motion of vehicles.

The graph for max-node degree in Figure 2 is obtained for three different road scenarios with changing safety distance between the cars such as 2m, 3m & 4m. Max-node degree graph is obtained by implementing dual-ring model for communication under same varying safety distances between cars.

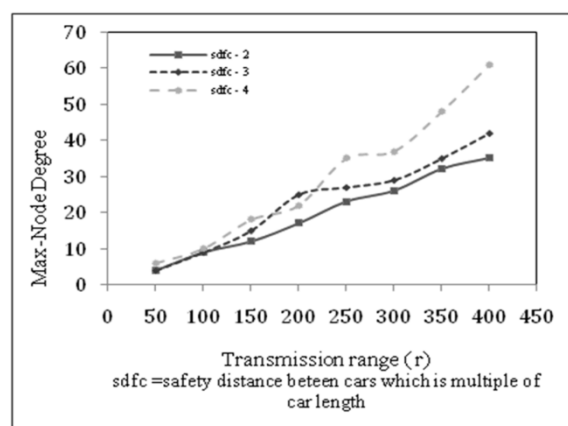


Figure 2: The network connectivity for cars and buses case having effect of varying safety distance in Kj on the dual lane

The jamming density is calculated by using the [19] having influence of heterogeneous road environment and changing road condition due to the front density. As the transmission range is increased, the graph for max-node degree shows different behaviour for different values of

safety distance. It is observed that after keeping a reasonable distance between the vehicles and implementing dual-ring communication for dual lane road infrastructure, the availability of backup path provides better conditions for the connectivity on the road as compare to single lane infrastructure. For the Figure 3 the arriving ratio for the car and bus is changed to observe the network behavior under the effect of changing heterogeneous traffic conditions on the roads. The four different road scenarios are created such that Probability of arriving cars (P_c) and Probability of arriving bus (P_b) is defined as $P_c / P_b = 12/08, 16/04, 08/12, 04/16$ for two lane road structure

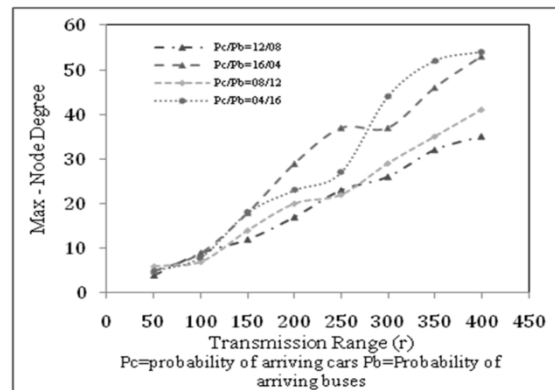


Figure 3: The network connectivity for cars and buses case having effect of safety distance in Kj with cars and buses in varying arriving probabilities on Dual lane under Dual Ring scheme

5. Conclusions

In this paper we have focused to get better connectivity dynamics in VANET under the effect of microscopic variables such as headway, safety distance and front density for heterogeneous traffic flow.

By considering the heterogeneousness in available type of vehicles new connectivity model is suggested. The movement of different kind of vehicles on the road impacts the traffic flow. With the use of desired type of urban traffic on the road for the communication network, we can improve connectivity and can get improvement in the traffic system. The implemented dual-ring model exhibit better connectivity conditions under the WEAC protocol structure.

The influence of important traffic parameters can be further analysed for the different road structures and conditions to overcome the issues of network structures in communication and road traffic congestion.

Acknowledgement

The authors are thankful to COMSATS Institute of Information Technology, Pakistan for their support for the advancement of research activities.

References

1. The Highway Code-<http://www.direct.gov.uk/> Accessed 10 Feb 2010.
2. S. P. Hoogendoorn and P. H. L. Bovy, "State-of- the-Art of Vehicular Traffic Flow Modeling," J. Sys. and Control Eng., Vol. 215, No. 4, 2001, pp. 283-304.
3. M. Artimy, W. Roberstson, and W. Philips, "Assignment of dynamic transmission range based on estimation of vehicle density," in Proc. of the 2nd ACM int'l workshop on Vehicular ad hoc networks (VANET'05), New York, NY, USA, 2005, pp. 40-48.

4. S. Panichpapiboon and W. Pattara-atikom, "Evaluation of a neighbour based vehicle density estimation scheme," in Proc. IEEE Int. Conf. ITST, Phuket, Thailand, Oct. 2008, pp. 294-298.
5. <http://www.tfsrc.gov/its/tft/tft.htm>. Accessed 2 January 2010.
6. M. Khabazian and M. K. Mehmet Ali, "A Performance Modeling of Connectivity in Vehicular Ad Hoc Networks," IEEE Trans. Veh. Technol, Vol. 57, No. 4, pp. 2440-2450, July 2008.
7. S. Yousefi, E. Altman, , R El-Azouzi, and M. Fathy "Analytical Model for Connectivity in Vehicular Ad Hoc Networks" IEEE Trans. Veh.Tech, Vol. 57, No. 6, pp. 3341-3356, Nov 2008.
8. S. Yousefi, E. Altman, R. El-Azouzi, and M. Fathy, "Improving connectivity in vehicular ad hoc networks: An analytical study," Comp. Commun., Vol. 31, No. 9, pp. 1653-1659, Jun. 2008.
9. Ivan W.H, Leung, Kin K, Polak, John W, Mangharam, Rahul. "Node Connectivity in Vehicular Ad Hoc Networks with Structured Mobility," 32nd IEEE Conference on Local Computer Networks, 2007, pp. 635-642.
10. M. Kafsi, P. Papadimitratos, O. Dousse, T. Alpcan, and J.-P. Hubaux, "VANET Connectivity Analysis," in Workshop on Automotive Networking and Applications (Autonet), 2008.
11. Ivan W. H. Ho, K. K. Leung, J. W. Polak, "Stochastic Model and Connectivity Dynamics for VANETs in Signalized Road Systems," IEEE/ACM Trans. on Networking (TON), Vol. 19, No. 1, Feb. 2011.
12. Umer, T, Zhiguo Ding, Honary, B, Ahmad, H, "Implementation of microscopic parameters for density estimation of heterogeneous traffic flow for VANET", in Proc. CSNDP, 2010, pp. 66-70.
13. K. K. Leung, W. A. Massey, and W. Whitt, "Traffic Models for Wireless Communication Networks," IEEE J. Sel. Areas. Commun, Vol. 12, No. 8, pp. 1353-1364, Oct. 1994.
14. W. A. Massey and W. Whitt, "Network of Infinite Server Queues with Non stationary Poisson Input," Queuing Syst., Vol. 13, pp.183-250, 1993.
15. W. A. Massey and W. Whitt, "A Stochastic Model to Capture Space and Time Dynamics in Wireless Communication Systems," Prob. Eng. Inf. Sci, Vol. 8, pp. 541-569, 1994.
16. T. R. Sheltami and H. T. Mouftah, "A Warning Energy Aware Clusterhead (WAC) for MANETs," Submitted to The IEEE Transactions on Wireless Communications.
17. C. J. Khisty and B. K. Lall, Transportation Engineering, Second Ed. Englewood Cliffs, NJ: Prentice Hall, 1998.
18. Ivan W. H. Ho, K. K. Leung, J. W. Polak, "Connectivity Dynamics for Vehicular Ad-Hoc Networks in Signalized Road Systems," in Proc. 21st International Teletraffic Congress (ITC), Sept 2009, pp.1-8.

Enhanced Positioning through Inference of Mobile Terminal Context

Mahsa Honary¹, Daniel Essafi², Lyudmila Mihaylova³

School of Computing and Communications, Lancaster University, Lancaster, UK¹

HW Communications Ltd, Lancaster, UK²

¹m.honary@lancaster.ac.uk, ²daniel@hwcomms.com, ³mila.mihaylova@lancaster.ac.uk

Abstract: This paper presents a framework for positioning in the London Underground system on a standard mobile phone. A rule based algorithm is applied that uses: (i) data provided by Transport for London (TfL) and (ii) readings from accelerometers on a standard android mobile phone. The system relies on a database built up from average times taken to complete transitions between adjacent underground stations. Results are presented showing the efficiency of the developed system.

Keywords: Positioning, Context Awareness, Mobile Phone, London Underground, Received Signal Strengths.

1. Introduction

For the vast majority of positioning applications the location result on its own is not useful. More specifically, providing a latitude and longitude does not provide much value to the end user. However, when the location is displayed on a map illustrating the relative locations of streets, buildings or the locations of other devices this information becomes useful.

Most current positioning applications tend to apply this ‘geo-tagging’ model, whereby relative locations can be viewed. In this research, this geo-tagging approach of directly using relative locations is extended and the value of Mobile Terminal (MT) Context is explored.

This can be best explained through the use of a simple example: If the RSSI values from neighbouring cells on an MT suddenly change, from the MT seeing several cells with strong RSSIs to the MT not seeing any cells it can be implied that:

1. The MT has moved;
2. The MT has moved from a location where signals were strong to somewhere where no signals can be observed (such as moving to a faraday cage);
3. The MT has moved to a location such as an underground building or lift.

With some additional information, such as knowing that the last location is close to a London Underground station, it can be inferred that the MT is probably inside the station, underground.

Not only does this contextual inference aid the ability to provide location reports, but also, it can provide information that can be used intelligently by the Location Sensors on the MT. For example, once it is inferred that the MT is in a particular underground station, it is known that certain Location Sensors will no longer work; and that other sensors should be activated. For instance on a smart phone, battery life can be saved by not trying to get a GPS fix, and by considering using information such as the accelerometer to deduce movement, direction and therefore current location.

This feedback loop between Location and Context is an interesting research topic, and the information that can be extracted from this loop has the potential to enrich information for the end user. This vision has been illustrated in the simplistic diagram shown below in Figure 1.

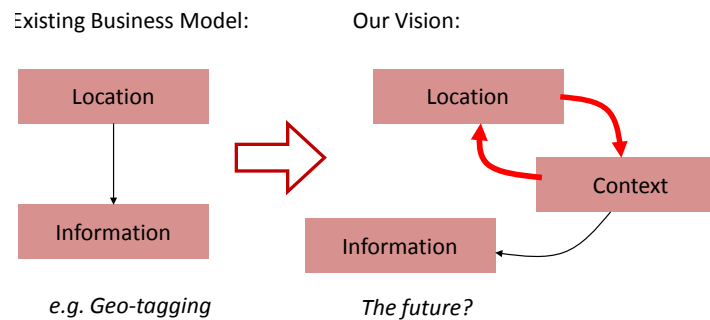


Figure 1: Location and Context feedback loop

In this paper the use case for positioning in the London underground system is considered. A rule base is applied that uses (i) data provided by TfL, and (ii) the readings from accelerometers on a standard android mobile phone.

The work considers the application where data may be required to be shown to a traveller at the right time, such as:

- The need to change routes midway in a journey;
- Notification as to how long it will take from where you are to where you want to go;
- Knowing which station you are in while underground, allowing for directions to be given to quickly navigate the station, i.e. turn left when you leave the train and look for the 'xyz' sign.

With this in mind the key goals for the positioning system can be stated to be:

- Knowing which station you are in (especially after having disembarked a train);
- Knowing which line you are travelling on;
- Knowing whereabouts on the line you are;
- Knowing whereabouts in a station you are.

2. Investigations

2.1. Initial View of the Problem

In the underground environment there are no GPS, WiFi and Cell signals or other fixed radio infrastructure to use for reference. This means that use must be made from the other sensors available in a Smartphone. In Android Gingerbread [1] there is the following selection:

- Accelerometer
- Gravity sensor
- Gyroscope
- Light sensor
- Linear Acceleration
- Magnetic Field sensor
- Orientation sensor
- Pressure sensor
- Proximity sensor
- Rotation Vector sensor
- Temperature sensor

On the face of it, Android appears to provide everything required need in a nice, neat package. Accelerations along the axes of the phone coupled with orientation values of yaw, pitch and roll should be enough to work out horizontal acceleration, which in this case should be the acceleration profile of a Tube train being travelled on. Accelerations in the horizontal plane are the basis for working the speed and displacement. However, there are additional

constraints which affect performance and scalability. More specifically, at present many phones are not equipped with a gyro sensor which is a prerequisite for working out horizontal acceleration. Therefore, the approach taken aimed to use only accelerometers which are available on all Android phones.

2.2. Methodology

The approach taken was to classify movement into 3 states: stationary, walking, possible train movement.

The classification was done by absolute levels of the standard deviations of resultant acceleration. The levels were tuned based upon data taken from people of varying weights/heights, on different phones, on different types of rolling stock (trains), on different lines including both automatically and manually driven trains. In addition a preliminary moving average filter was applied to remove large fluctuations from the data.

The rule base for classification applies Mamdani trapezoid membership functions [2], for the standard deviations (as mentioned above) as well as the time taken to travel between adjacent London Underground stations. Standard 'if', 'then' rules were coded to enable the computation of the rule base in near real-time on a lightweight mobile phone. In this application, the parameters for the rule base have been formed based on previously collected data.

TfL publish various data feeds for the London Underground. These are web services that return data in JSON format [3]. They give live data every 30 seconds. The feed of interest is the 'Tube Departure Boards' feed. This states where each train on a line is at the time it updates. So every 30 seconds we are told that there are e.g. 5 trains on the northern line bank branch travelling north towards High Barnet. Each train has an ID and an estimated time of arrival to the next station.

The method applied was to poll and log all the available information as quickly as possible, in our case every 30 seconds. This enabled us to follow each train and identify how long they take to move between stations. Often it was not possible to identify trains moving from one station to the next, but it was possible to infer times between adjacent stations from different trains.



Figure 2: Example of London Underground Station Network

If we consider the simplistic example of the London Underground station network illustrated in Figure 2. If we want the times from B to C, but it is not in the data set, we can infer the time from the time taken from A to C and the time taken from B to D.

This data set was sampled continuously for around two weeks and the results were analysed statistically. In addition, travel times were benchmarked against physical distance as a sanity check. The end result was a dataset containing all adjacent pairs of stations, the distance between them, and the times a train actually took on that leg.

This data was then utilised as the 'Tube Network Data' membership function. Current location was the inferred by using this membership function and the membership functions associated with the movement classification (using the standard deviations from the observed accelerations from the phones).

In matching accelerations to the Tube Network Data we had to take unscheduled stops into account. Tube trains frequently stop in tunnels and do not always travel between stations at the same rate. To accommodate this behaviour we had to match actual travel time between stations to the Tube Network Data as calculated by our state engine.

It should be noted that the approach taken is designed to be applied for real-time calculation of location on a phone in the underground. This important constraint led to an approach tailored to speed of calculation and robustness of immediate result.

3. Observations

The resultant acceleration for a short journey is presented in Figure 3.

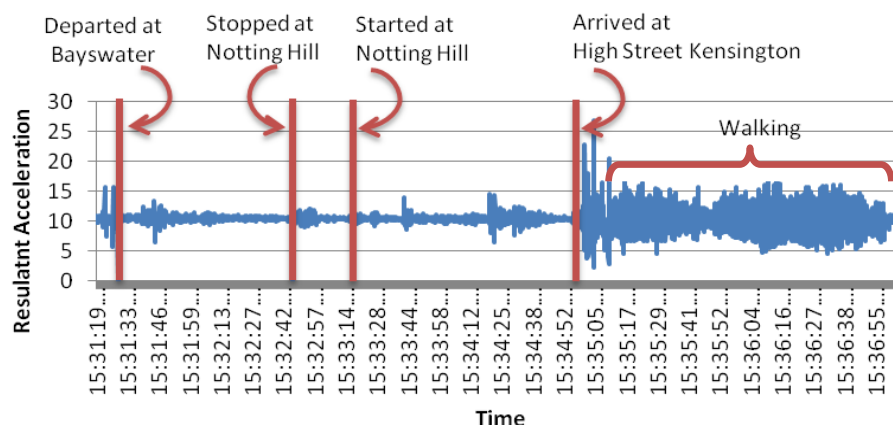


Figure 3: Resultant accelerations observed for example journey on London Underground

Figure 3 clearly shows that the standard deviation for walking is significantly greater than that observed when travelling on the London Underground. Furthermore, when stationary the standard deviation of the resultant acceleration is significantly lower than when travelling on the London Underground. However, there are challenges when travelling on an escalator, where the patterns of the fluctuations of the resultant acceleration are very similar to that observed when travelling on the London Underground trains.

Distinguishing between movement on an escalator and the underground train could not be completed using resultant acceleration alone. To distinguish an escalator event consideration was given to the time between walking events; and the discrepancy of this time with that expected for the train to travel between adjacent underground stations.

4. Results

The data set used for testing was selected to cover all lines, all types of rolling stock (trains), both automatically driven and manually driven. Several devices were used to capture and later analyse the effect of different accelerometer hardware and firmware. Further non-tube data was collected to study the accelerations associated with different patterns of walking for people of different weights, different ways of carrying a phone, effect of escalators and elevators. The data comprises:

1. 11 journeys comprising a total of 39 journey segments with data collected on HTC Nexus One [a journey segment is a tube journey from one station to the next station along a line, it comprises both a start and a stop].
2. 9 journeys comprising a total of 19 journey segments with data collected on HTC Nexus One.

3. 10 journeys comprising a total of 67 journey segments with data collected on Samsung Nexus S.
4. 2 journeys comprising a total of 4 journey segments with data collected on HTC Nexus One.
5. 12 journeys comprising a total of 24 journey segments with data collected on Samsung Galaxy S2.
6. 2 journeys comprising a total of 4 journey segments with data collected on HTC Wildfire.

Designing and optimizing the algorithm over the whole set of data opens the door to problems of over-fitting. To avoid this problem the data was randomly split into 2 sets, one set for use in design & optimization and the other for testing for over-fitting. This process was performed many times with different randomly selected data each time. Once this iterative process was complete and the algorithm fixed, it was run over the whole data set.

The quality or performance of the algorithm is assessed according to how well it catches starts and stops in the data. We tested against two measures: % of location estimates within 30 seconds of the actual real-world event and % within 60s. Where a location is not returned, this was counted as outside 60 seconds. For the whole data set (more than 300 location events) the results are:

Within 30 seconds = 81%
Within 60 seconds = 87%

Using the same number-bullet points above, running the algorithm on the subsets of data described gives the following results:

1. Nexus One, 82% inside 30s & 89% inside 60s
2. Nexus One, 75% inside 30s & 86% inside 60s
3. Nexus S, 80% inside 30s & 83% inside 60s
4. Nexus One, 100% inside 30s & 100% inside 60s
5. Galaxy S2, 85% inside 30s & 93% inside 60s
6. HTC Wildfire, 90% inside 30s & 90% inside 60s

The following plots (Figures 4a, 4b and 4c) illustrate the variables used to classify movement. The following key applies:

'oMarker' shows the actual events where 10 represents movement on the underground train and -10 illustrates no movement (stationary)

'Res' shows the resultant acceleration (based on the x axis on the left hand side of the plots)

't3' shows the calculated 'state' of the traveller, 0 = stationary, 1 = possible train movement, 2 = walking/running

'dTime' shows the time since last detection of Walking movement used to help identify escalators

'event' shows the output of the location reporting engine, whereby 10 represents movement on the underground train and -10 illustrates no movement (stationary)

'Var' shows the variance (or a function thereof) of the resultant acceleration

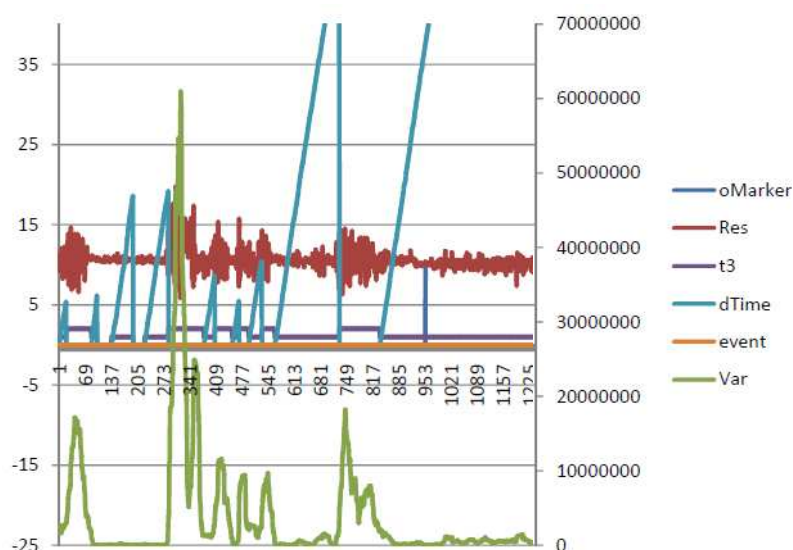


Figure 4a: Movement classification parameters

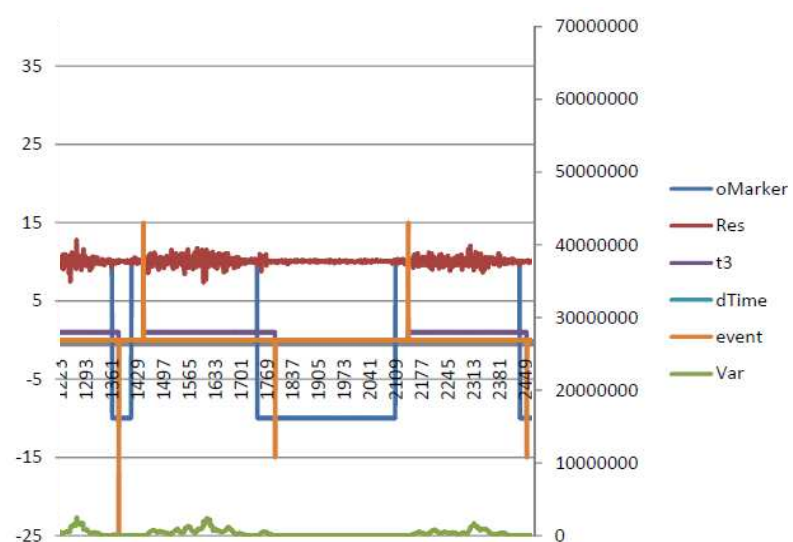


Figure 4b: Movement classification parameters

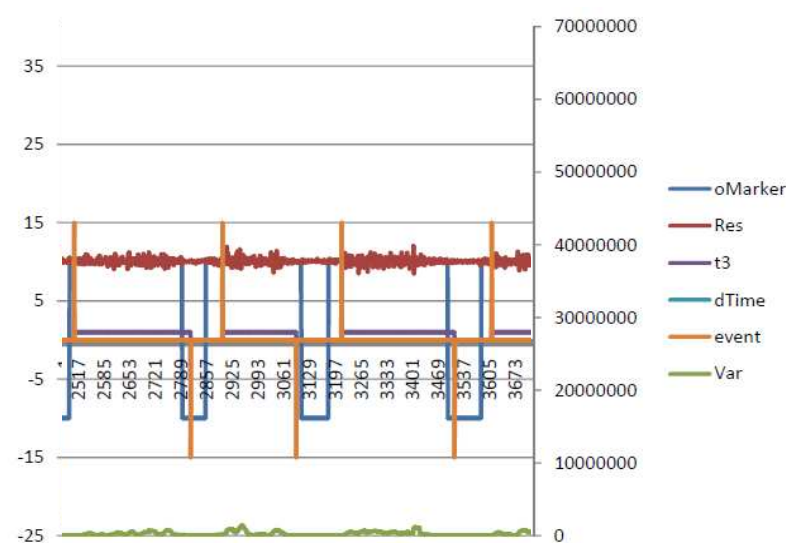


Figure 4c: Movement classification parameters

5. Conclusions

The research shows that 87% of the time the location can be returned inside the London Underground network using only the starting location (taken from outside the station), the accelerometer and data associated with the time taken between adjacent stations.

The approach taken applies the context that the user has entered the London Underground station based upon the sudden loss of RF signals, and the fact that the last known location was near to the Underground station. This context is then used to activate the movement classification described in the paper from which user location is inferred.

Further work will be focussed on using the patterns of the resultant accelerometer to help define the speed at which the train is moving, and in turn identify the distance travelled. This may help overcome some of the issues observed, and increase accuracy above the current 87%.

Acknowledgement

The authors would like to acknowledge the support of HW Communication Ltd and EPSRC for the support they have provided in enabling this research to be undertaken.

References

1. Android 2.3 “Gingerbread” Platform API documentation,
<http://developer.android.com/sdk/android-2.3.html>
2. Mamdani, E. H. and S. Assilian, “An experiment in linguistic synthesis with a fuzzy logic controller,” Int. J. Man-machine Studies, Vol. 7, pp. 1–13, 1975.
3. Transport for London live travel information feed,
<http://www.tfl.gov.uk/businessandpartners/syndication/16493.asp>