Volume 36 Number 2 June 2012 ISSN 0350-5596

Informatica

An International Journal of Computing and Informatics

Special Issue:

Human Being in the Digital World

Guest Editors:

V. Fomichov

O. Fomichova

1977

Editorial Boards, Publishing Council

Informatica is a journal primarily covering intelligent systems in the European computer science, informatics and cognitive community; scientific and educational as well as technical, commercial and industrial. Its basic aim is to enhance communications between different European structures on the basis of equal rights and international refereeing. It publishes scientific papers accepted by at least two referees outside the author's country. In addition, it contains information about conferences, opinions, critical examinations of existing publications and news. Finally, major practical achievements and innovations in the computer and information industry are presented through commercial publications as well as through independent evaluations.

Editing and refereeing are distributed. Each editor from the Editorial Board can conduct the refereeing process by appointing two new referees or referees from the Board of Referees or Editorial Board. Referees should not be from the author's country. If new referees are appointed, their names will appear in the list of referees. Each paper bears the name of the editor who appointed the referees. Each editor can propose new members for the Editorial Board or referees. Editors and referees inactive for a longer period can be automatically replaced. Changes in the Editorial Board are confirmed by the Executive Editors.

The coordination necessary is made through the Executive Editors who examine the reviews, sort the accepted articles and maintain appropriate international distribution. The Executive Board is appointed by the Society Informatika. Informatica is partially supported by the Slovenian Ministry of Higher Education, Science and Technology.

Each author is guaranteed to receive the reviews of his article. When accepted, publication in Informatica is guaranteed in less than one year after the Executive Editors receive the corrected version of the article.

Editors's Introduction to the Special Issue on "The Human Being in the Digital World"

The stormy progress of information and communication technologies (ICT) since the end of the 1980s has wonderfully changed not only professional but also every-day life of very many people in the world. Now it is possible to speak with a friend across an ocean and to see your friend; due to cell telephones, the mothers are able to immediately find out what and where their children are doing; we do have the possibility to obtain in several minutes the e-versions of thick books stored thousands miles away from our terminals, etc.

However, during last decade many scholars from different countries have observed and discussed a number of negative tendencies in the development of the personality and society caused by the expansion of ICT and globalization processes. The authors of the papers presented below look for constructive ways of contributing to harmonic development of the personality in modern information society. The common feature of the papers is that they either relate to the new scientific discipline called Cognitonics or correspond to its goals.

Cognitonics is a new scientific discipline emerged in the first half of the 2000s (its prenatal stage of development covers the years 1993 – 2003). It aims (a) at explicating the distortions in the perception of the world caused by the information society and globalization and (b) at coping with these distortions in different fields by means of elaborating systemic solutions for compensating the negative implications of the kind for the personality and society, in particular, for creating cognitive-cultural preconditions of the harmonic development of the personality in the information society and for ensuring the successful development of national cultures and national languages.

The birth of Cognitonics was stimulated by the ideas of Philosophy, Cognitive Linguistics, Artificial Intelligence theory, Web Science, Applied Linguistics, Art theory, Cognitive Psychology, and Cognitive Biology.

Two factors seem to be especially important from the standpoint of achieving the goals of Cognitonics:

- information and communication technologies have been developing extremely quickly and have been expanding unusually broadly, they penetrate not only into every office and laboratory but also into every school class and every family;
- it is necessary and promising to use the power of modern ICT in order to very quickly and broadly disseminate the elaborated effective methods of compensating the negative distortions in the development of the personality and of national cultures in information society.

From the standpoint of educational practice, Cognitonics proposes an answer to the following question: what precious ideas and images accumulated by the mankind, at what age, and in what a way are to be inscribed into the conceptual picture of the world of a person in order to harmonize his/her intellectual and spiritually-coloured emotional development and to contribute to the successful development of national cultures and national languages.

Overview of the issue

This special issue of Informatica – an International Journal of Computing and Informatics contains 6 papers submitted by the researchers from two continents (Europe and North America) and 6 countries: Croatia, Italy, Mexico, Poland, Romania, and Russia. The papers were carefully selected on the basis of peer reviews. The authors of four from six papers use the advantages of modern ICT for constructively contributing to the harmonic development of the personality in modern information society.

The title of the first paper is "A Contribution of Cognitonics to Secure Living in Information Society" (V.A. Fomichov and O.S. Fomichova, Russia). The paper sets forth the weighty arguments in favour of much earlier socialization of children in the Internet age. The first goal of the described study is to make children (including teenagers) be aware of possible social consequences of their misuse of ICT, in particular, of the cell telephones and the Internet. The second goal is to find a way of inscribing into the conceptual picture of the young child of the respect to the ideas formulated by another person's. An original approach to early forming the cognitive subspace of moral values and social responsibility is proposed. It is a part of the System of Emotional-Imaginative Teaching (the EIT-system) developed and tested by the authors during 1990s – 2000s. For describing this method, a new formal notation for representing transformations of the learners' cognitive-emotional sphere and the spectrum of information processing skills is proposed, it is called the notation of the maps of cognitive transformations. The described method of early socialization and the EIT-system as a whole are interpreted as an important component of Cognitonics. The next subject of the paper is a new way of considering impressionism under the frame of Cognitonics. An original algorithm of transforming the negative emotions (caused by the messages received from social networks) into the positive ones is proposed. This algorithm considers the possible reactions of a human (including the recommended reactions) to the emotional attacks via social networks. It is proposed to include an analysis of the kind into the program of the course "Foundations of secure living in information society".

The subject of the paper "Information Technology for Management and Promotion of Sustainable Cultural Tourism" (M. Valčić and L. Domšić, Croatia) is the use of intelligent ICT solutions in the development of heritage tourism. The aim is to contribute to preserving national culture, creating partnership and enhancing destinations value in information society. The authors propose a model for local Destination Management Systems of heritage destinations (DMS), the purpose is to help to achieve a more globally responsible paradigm for the tourism industry and facilitate the management of destinations and the coordination of the local suppliers. DMSs provide interactive demonstrations of local amenities and attractions and enable consumers to build their own itinerary based on their interests and requirements. Under the framework of this approach, all the stakeholders within the destination are linked with each other in order to create collaborative action and a genuine, sustained growth in heritage tourism. The paper analyzes, as an example, the Croatian World Heritages Sites and their status on the Web.

The paper "Linguistic Model Propositions for Poetry Retrieval in Web Search" (M. Granos and A. Zgrzywa, Poland) excellently corresponds to a new, large-scale goal formulated by Cognitonics for the software industry and Web science. This goal is to develop a new generation of culture-oriented computer programs and online courses (in the collaboration with educators, linguists, art historians, psychologists) - the computer programs and online courses intended for supporting and developing positively-oriented creativity, cognitive-emotional sphere (in other words, emotional intelligence), the appreciation of the roots of the national cultures, the awareness of the integrity of the cultural space in the information society, and for supporting and developing symbolic information processing and linguistic skills, associative and reasoning abilities of children and university students.

The paper grounds the importance of constructing the advanced exploratory search engines, which, according to the Maslow's hierarchical pyramid of needs, address the needs of higher order associated with self-fulfillment and satisfaction of informational-cognitive aesthetic ambitions of non-verbal transmission. A number of linguistic models for poetry retrieval purposes are studied.

The paper "The Experiences of Landscape Social Per-ception as a Remedy for Plunging into Virtual Reality" (R. Micarelli and G. Pizziolo, Italy) proposes a possible answer to the question "How to prevent young people from plunging into virtual reality"? It is a very socially significant question, because numerous observations in different countries have shown that plunging into the virtual life prevents people (especially young people) from solving real life problems, from establishing social relations with the people around them, and often causes breaking family ties and other social ties. Usually, a person being carried away by virtual reality is unsuccessful in solving the problems of real life (if some external circumstances force he/she to do it).

The paper "Using M Tree Data Structure as Unsupervised Classification Method" (M.C. Mihăescu and D.D. Burdescu, Romania) describes a new theoretical and practical approach to the problem of automatically providing information to the students and course managers regarding the knowledge level reached by the students. The purpose is increasing the effectiveness of educational processes. The proposed approach is based on the usage of M Tree structure for classification of the learners based on their final marks obtained in their respective courses. The classical building algorithm of M-Trees with an original accustomed clustering procedure was implemented. The data that are managed within M Tree structure are represented by the instances. A baseline classification scheme based on k-means clustering and a custom M Tree clustering are presented. For comparison, the classical characterization formulas are considered.

The subject of the paper "Computer-Aided Educational Intervention in Teenagers Via Internet Social Networking" (M. G. Velázquez-Guzman and F. Lara-Rosano, México) is the experience of contributing to forming the new citizens but not only to teaching mathematics, languages, and other disciplines. The authors of the paper underline that new citizens should be critical and creative social actors participating in the construction of better ways of coexistence. As a possible way of achieving this goal, the paper proposes educational strategies based on the use of Internet social networking sites as Facebook®, Twitter® and MySpace®. These described strategies must enhance the social development of the students, taking into account their peers' subculture.

The guest editors would like to thank the Managing Editor of Informatica for the kind invitation to prepare this special issue on The Human Being in the Digital World. Finally, many thanks to the authors of the papers for their contributions and to all of the referees for their precious comments ensuring the high quality of the accepted papers and making the reading as well the editing of this special issue a rewarding activity.

Vladimir A. Fomichov, Olga S. Fomichova
Guest Editors

A Contribution of Cognitonics to Secure Living in Information Society

Vladimir A. Fomichov
Faculty of Business Informatics, National Research University Higher School of Economics
Kirpichnaya str. 33, 105187 Moscow, Russia
E-mail: vfomichov@hse.ru, vfomichov@gmail.com

Olga S. Fomichova
State Educational Centre "Dialogue of Sciences", Universitetsky prospect 5, 119296 Moscow, Russia

Keywords: early socialization of children, informational-aesthetic conception of developing cognitive-emotional sphere, maps of cognitive transformations, theory of dynamic conceptual mappings, methods of emotional-imaginative teaching, algorithm of resisting emotional attacks via social networks, impressionism

Received: December 15, 2011

The paper grounds the necessity of much earlier socialization of children in the Internet age. The main goal is to make children (including teenagers) be aware of possible social consequences of their misuse of information and communication technologies, in particular, of the cell telephones and the Internet. An original method of early forming the cognitive subspace of moral values and social responsibility is stated. It is a part of the System of Emotional-Imaginative Teaching (the EIT-system) developed and tested by the authors during 1990s – 2000s. For describing this method, a new formal notation for representing transformations of the learners' cognitive-emotional sphere and the spectrum of information processing skills is proposed, it is called the notation of the maps of cognitive transformations. The described method of early socialization and the EIT-system as a whole are interpreted as an important component of cognitonics - a new scientific discipline. The paper also represents a new way of considering impressionism under the frame of cognitonics. An original algorithm of transforming the negative emotions (caused by the messages received from social networks) into the positive ones is proposed. This algorithm considers the possible reactions of a human (including the recommended reactions) to the emotional attacks via social networks. It is proposed to include an analysis of the kind into the program of the course "Foundations of secure living in information society".

Povzetek: Prispevek se ukvarja z zgodnjo socializacijo otrok pri uporabi spleta.

1 Introduction

The stormy progress of the Internet since the 1990s has tremendously expanded the sphere of using the informational technologies (IT). In the developed countries, IT have reached practically every home. One of the consequences is that IT have considerably speeded-up the globalization processes. Every coin has two sides. During two decades the scholars in various countries have observed a number of negative shifts in the development of the personality, national cultures and languages caused by IT and globalization processes. These were the principal reasons for the birth of a new scientific discipline – cognitonics [7 - 10, 13 - 15]. It aims (a) at explicating the distortions in the development of the personality and national cultures caused by the peculiarities of information society and globalization and (b) at coping with these distortions in different fields by means of elaborating systemic solutions for compensating the negative implications for the personality and society of the stormy development of informational technologies and globalization processes, in particular, for creating cognitive-cultural preconditions of the harmonic development of the personality in the information society and for ensuring the successive development of national cultures and national languages.

From the standpoint of educational practice, cognitonics proposes an answer to the following question: what precious ideas and images accumulated by the mankind, at what age, and in what a way are to be inscribed into the conceptual picture of the world of a person in order to harmonize his/her intellectual and spiritually-coloured emotional development and to contribute to the successful development of national cultures and national languages.

One of the serious, large-scale negative phenomena observed mainly during last decade is that the teenagers (and some times younger children) have received the possibility to distribute any information about their peers and adults with the help of the cell telephones and the

Internet. Unfortunately, a small part of children have used this possibility for bullying, in particular, for distributing discreditable photographs and texts [17].

That is why we believe that a possible way out is to elaborate the methods of much earlier socialization of the child than it is usually done in order, on the one hand, to eliminate or considerably diminish children's aggressiveness. On the other hand, for contributing to the birth in children of the feeling of social responsibility and to the understanding by children of the severe consequences suffered by their peers and adults.

The arguments of the kind form the content of Section 2 of this paper. Section 3 describes the central ideas of an original informational-aesthetic conception of developing the cognitive-emotional sphere of the learners: young children, teenagers, and university students. These ideas are realized by our System of the Methods of Emotional-Imaginative Teaching (the EIT-system), it indicates a new way of solving the problem of much earlier socialization of children in the computer age. Section 4 sets forth the objectives and structure of the EIT-methods. Sections 6 - 7 outline the principal ideas of our approach to solving the problem of much earlier socialization of children. For this, a new graph notation is introduced in Section 5 – the notation of the maps of cognitive transformations (MCT). Each MCT describes the transformations of the cognitive-emotional sphere of the learner and of the spectrum of his/her cognitive skills achieved as a result of employing pedagogical methods. The subject of the sections 8 – 13 is an original algorithm of resisting emotional attacks realized by means of social nets. It is based on the acquaintance of the learners with the principal ideas of impressionism. Section 14 contains the conclusions.

2 The Need of Much Earlier Socialization of Children in Information Society

Let's consider a number of phenomena caused by IT that can be interpreted as negative shifts in the development of the personality in information society. A considerable part of information distributed via Internet is false, but this false information is "injected" into the net by concrete people. For instance, in March of 2011 a blacksmith from the island Crete posted the following anonymous e-mail message: "Bankruptcy is a matter of days and has been scheduled for the 25th of the month". This message was reproduced by dozens of Greek blogs and caused a kind of national panic [2].

Very many high school students and university students don't feel the value of a thought generated by another person and, as a consequence, use in their works the ideas, methods, models belonging to another people. Acting in this way is very easy, because the users of the Internet have access to a huge amount of informational sources physically stored far away from the user.

Twenty or more years ago the typical consequences of a conflict between a child and his/her classmates were the use of insulting nicknames or (in some schools and some classes) the fights between the classmates. It was bad for a child involved in a conflict of the kind, but nowadays the consequences of such conflicts may be even tragic. The reason is that now the school students of even middle grades possess the well-developed skills of using e-mail, Internet, cell telephones. However, they are not socially mature and are not ready to suffer all the consequences of their deeds.

The paper [17] describes a new, tragic phenomenon of information society: cyberbullying. According to that paper, cell telephones, profile home pages and social networking service are a breeding-ground for cyber-bullying in modern Japan. The moral pressure imposed due to these technical means may become unbearable for the teenager, and, unfortunately, there are known the cases when the child decides to leave the life. The other negative deeds performed with the help of modern information and communication technologies are the attacks of young hackers against computer systems of socially very significant objects and military objects.

The common reason for all considered negative deeds is that children possess very high skills of using IT but are very far from being socially mature and, as a consequence, from being aware of own social responsibility. An analogy can help to grasp the essence of such situations. A tiger-cub, while playing with a chicken, hit the chicken with his paw. The tiger-cub was not aggressive, he didn't attack the chicken, didn't want to hurt the chicken on purpose. It was just playing, the tiger-cub was not aware of the power of his paw in comparison with the weakness of the chicken, but, as a consequence, the chicken couldn't move any more.

The same happens to young children and teenagers, they can't clearly understand the severity of their actions many times multiplied by the power of modern IT. A simple negative intention is turning into a powerful tool of destroying the human's Self. And, in return, the Self of the one who becomes aware of his/her power is in danger too: he/she is not ready to suffer the consequences.

3 About an Informational-Aesthetic Conception of Developing Cognitive-Emotional Sphere of the Learners

The educational methods stated in this paper belong to the System of the Methods of Emotional-Imaginative Teaching (EIT-system). The core of the EIT-system was elaborated by O.S. Fomichova in the first half of the 1990s and has been expanded in the second half of the 1990s and in the 2000s. This system is underpinned by our Theory of Dynamic Conceptual Mappings (the DCM-theory). This theory is stated in numerous publications both in English and Russian, starting from the paper [3]. Both the DCM-theory and the EIT-methods form a principal part of the cognitonics constructive core.

A main component of the DCM-theory is an original informational-aesthetic conception of developing the cognitive-emotional sphere of the learners: young children, teenagers, and university students. The central ideas of our conception are as follows.

1. It is important to actively develop a broad spectrum of information processing skills of the child, starting at least at the age of five. It applies, in particular, to associative abilities, the skill of integrating information from various sources, and the ability of establishing time-causal relationships between the events (see [4, 5]).

2. It is very important to combine the development of information processing skills with inscribing, in a systemic way, the feeling of beauty into the world's conceptual picture of the child. Proceeding from our experience accumulated during 21 years, we consider the following educational processes as the principal instruments of achieving this goal:

- early support and development of figurative (or metaphoric) reasoning;
- teaching young children (at the age of 5 – 6) very beautiful language constructions for expressing the impressions from the nature;
- a unified symbolic approach to teaching natural language (mother tongue and a foreign language), the language of painting and the language of dance [4, 10, 14].

3. Passing ahead the development of soul in comparison with the development of reasoning skills. A well-developed feeling of beauty plays an especially significant role in the realization of this idea. Besides, it is very important to be aware of the fact that children should have enough time for the development of soul: the time for contemplation, for imbibing the beauty of the nature, etc., i.e. children should have time for self-paced activity [10].

4. The principal cognitive precondition of successful (as concerns a long-term perspective) acquainting children with computer is the realization of the Thought-Producing Self of the child. It means that the child should know that his/her thoughts may have a high social significance, that is, be appreciated by his/her peers, by parents, grandparents, the teacher, etc. (see [6, 7, 10, 12]). The child should be aware of this fact before the time when the adults start to systematically acquaint him/her with computer.

5. Due to mastering modern information and communication technologies (ICT): cell telephones, internet, etc., the consequences of children's negative actions may be very severe. That is why it is necessary to find the ways of much earlier socialization of children in the modern information society in order to eliminate or considerably diminish their aggressiveness and to contribute to realizing by children the real scale of their misuse of ICT.

4 Shortly About the System of Emotional-Imaginative Teaching

The principal goal of the EIT-system is to develop in young children and teenagers:

- the skills of processing symbolic information, the reasoning abilities;
- a mature, rich cognitive-emotional sphere, the ability to perceive and appreciate the beauty in all its manifestations, in particular, in the deeds of people [14];
- the ability to understand the peculiarities of the conceptual picture of the world of the communication partner;
- the understanding of the complex system of social agreements;
- the skill of grounding the own point of view, of participating in a dialogue;
- the feeling of belonging to a very long chain of previous and future generations as the principal cognitive precondition of sustainable development;
- proud as concerns the own connection with great national culture, the openess to the achievements both of national and world culture.

For achieving the indicated objectives, a collection of the interrelated educational methods and an original cross-disciplinary educational program have been developed by O.S. Fomichova. The elaborated program is intended for teaching children during twelve years, where the starting age is five to six years. The program has been personally tested in Moscow with great success by O.S. Fomichova over a period of 21 years. It includes the following series of lessons: (1) a two-year course (the age of learners is 5 to 7 or 6 to 8 years) of studying foundations of reading and speaking English as a foreign language (FL), including learning basic elements of English grammar (Present Simple and Past Simple Tenses); (2) a course on understanding the language (a part of FL) of describing the nature and feelings evoked by nature; (3) a course on understanding the symbolic language of painting; (4) a course on understanding the language of poetry (with the accent on understanding metaphors and descriptions of nature); (5) a course aimed at (a) first acquaintance with sciences and (b) developing the abilities to argument the own opinion, to raise objections, etc.; (6) a course on improving the knowledge of English grammar (during mainly the fifth year of studies); (7) the course "Foundations of secure living in information society". In fact, the lessons of courses (2) to (7) may interchange [10].

The developed program can be interpreted as a model of a system of disciplines traditionally learned in school (elementary, middle, high) and in the courses forming the humanitarian component of university education. The main thing is that this program incorporates the interacting elements belonging to different disciplines and providing the possibility to achieve in practice the goals of cognitonics as concerns the well-balanced development of the personality in information society.

5 The Maps of Cognitive Transformations

During last two decades, at least two graph notations have been introduced and employed for explicating the essence of educational processes. The vertices (or nodes) of a concept map represent the basic concepts of a studied discipline, and the edges (or arcs) correspond to the relationships between these concepts [1, 16]. The conceptual-visual dynamic schemes (CVD-schemes) are the marked oriented graphs introduced by V.A. Fomichov and O.S. Fomichova, in particular, in [3, 7, 11] for inventing effective teaching analogies. Such graphs establish a correspondence between the components of a piece of theoretical material to be studied and the components of a well-known or just created by the teacher but bright fragment of the inner world's picture of the learner.

One of the principal goals of cognitonics is to create the cognitive-emotional preconditions of the well-balanced, harmonic development of the personality in information society. That is why we propose a new graph notation allowing for reflecting the initial and achieved states of the cognitive-emotional sphere of the learner.

By definition, *a map of cognitive transformations (MCT)* is an oriented graph with the vertices of three classes (or types). The *A-vertices* are represented by the rectangles (or blocks) with *single contour*; the texts inside these blocks describe the theoretical materials underpinning the teaching methods. The rectangles corresponding to *B-vertices* have the *double contour;* the marks inside these rectangles are the texts describing the activity at a lesson (or lessons) either of a teacher or of the students. The *C-vertices* are represented not by the rectangles but *by the ovals;* the texts inside these ovals describe the initial or achieved state of the cognitive-emotional sphere of the learner or the state of the spectrum of the learner's cognitive skills.

The Figures 1 - 3 commented in the next sections are the examples of the maps of cognitive transformations.

6 The First Stage of Supporting and Developing Reasoning Skills and Creativity of the Child

The foundation of educational activities aimed at achieving the objectives of our informational-aesthetic conception of developing the cognitive-emotional sphere of the learners is the first stage of supporting and developing the reasoning skills and creativity of the child. A map of cognitive transformations realized at this stage is presented on Figure 1.

One of the distinguishing features of our approach to this problem is that it is realized at lessons of a foreign language (FL) – English, where the mother tongue of children is Russian. The use of original, specially invented analogies (being the parts of fairy-tales and thrilling stories) for teaching the English alphabet, the rules of reasoning, and the basic rules of English grammar contributes to developing associative abilities of children at the age of 5 – 6.

Example. A difficult problem is to explain to very young children why the verbs in the 3rd person of Past Simple Tense have no ending "s", but the same verbs in the 3rd person of Present Simple Tense do have such ending ("reads" but "read", etc.). An interesting story from one of the previous lessons associates in the consciousness of the child the ending "s" with a bow. That story about Mr. Do and Lady Teacher is given in [11]. The teacher explains that her young students were in the Past babies and had no hair (were bald). Hence it was impossible to tie a bow. That is why verbs have no ending "s" in the 3rd person of Past Simple Tense. The 5-year-old students accept this explanation with great joy and remember it very well. As a result of having heard the stories of the kind, young children become aware of the fact that symbolic objects have the meanings pertaining to the real or fairy-tale life.

The interesting stories about the life of verbs and other words establish in the consciousness of the young child a mapping from the objects and situations of the real life to the domain of language entities (verbs, nouns, pronouns, etc.). That is why the consciousness of the young child becomes a considerable impulse to developing the ability to establish diverse analogies.

The other reason for using the lessons of FL is that (as a 21-year-long experience has shown) young children easier learn beautiful language constructions for describing the impressions from the nature than the equivalent constructions in mother tongue (see [4]). The explanation of this phenomenon is that in the first case children don't feel any contradiction with the every-day use of language.

Example. Let's consider a fragment from the home composition "The Winter Day", it was written in English by a six-year-old Russian speaking student Kseniya of the second year of studies in experimental groups:

"In the picture I see a winter day. On the branches of fir-trees, pine-trees, and birches lies fluffy, white snow, glimmering in sunshine. It seems that snowdrops are covered with jewels. Near the wood there are fields with snow. On the edge of the snowfield the rill is dreaming. The snow is everywhere. Sunrays make one way through the grey, big, heavy clouds and run over top-trees. Pine-trees and fir-trees shine hoary green. The bare bushes of birches are covered with snow. It seems that the oak is with soft, white, and big leaves.

Suddenly someone in heaven has dashed a big cup of sunlight upon the Earth, and the big old oak has turned into a fairy King in orange magnificent gown. And around him young birches in nice gowns are accompanying their beloved King".

7 A New Method of Forming the Cognitive Subspace of Moral Values and Social Responsibility

Taking into account the existence of the mentioned negative phenomena, the following new objectives for influencing the development of the personality in modern information society can be formulated.

It is necessary to start earlier than it is traditionally done to acquaint children with the very complex system of social agreements. Since this system is based on numerous symbols, the scholars need to pay more attention to developing symbolic information processing skills of young children and teenagers. In addition, it is necessary to early acquaint children with the idea that different people may have considerably different inner world's pictures (i.e. conceptual systems), and it is very important to take this into account while interacting with people. It is important to explain to young children and teenagers that practically every person has various connections with many other people. That is why a suffer of a classmate, etc., in fact, causes the suffer of many other people: mother, father, brothers, sisters, grand mother, grand father, etc.

The Figures 2 and 3 represent the basic components of our method. The central idea of the method is as follows. Rather often a child tries to distinguish himself/herself among his/her peers by means of emulating a bad pattern of the adults' behaviour: smoking, aggressiveness, following the formula "Might is right", etc. This applies not only to the teenagers but also to children at the age of 10 – 11. It is important to underline that the bad patterns of the adults' behaviour (drinking, etc.) most often are the consequence of misfortune, despair. Normally, children have no despair, they simply emulate the adults.

As a result of employing the stated method at lessons under the framework of the EIT-system, the child acquires (by the end of the second year of studies, at the age of 6 – 7 years) the possibility to distinguish himself/herself not by means of a deviant behaviour but with the help of mature thoughts and thoughtful behaviour.

The figure 3 shows that one of the important preconditions of employing our method during the second year of studies is the well developed figurative (or metaphoric) thinking. The scheme of creating this precondition is as follows:

Reading and discussing complex texts in English as a foreign language (FL) at the age of 5 –6

→ mastering a rich sublanguage of FL for expressing the beauty of nature and the feelings evoked by nature

→ development of figurative reasoning + development of the awareness of the social role of Natural Language

→ understanding poetical metaphors

→ creating metaphors

→ understanding the symbolic language of painting

→ development of the ability of decoding the messages conveyed by the masterpieces

→ realization of the "Thought-Producing Self" (see [6, 7, 10, 12]) and the improvement of the feeling that a person is a link in the long chain of previous and future generations.

8 The Role of Studying Impressionism

According to the dictionary, an illusion is a false idea, belief, or impression. One of the greatest illusions created by the Internet is the illusion of true life in the Cyberspace. One can earn money, spend free time, enjoy communication, go shopping, and even enjoy evil delight making other people harm. But an unpleasant side is that the Cyberspace lets teenagers pretend, lead an imaginary life under another imaginary name (nick).

Impression, the first impression in particular, is an ability of human intelligence. It is an idea, a feeling, or an opinion about something/somebody, especially one that is formed without conscious thought. The first impression is very often far from being true. It is caused by a detail and trifle.

Though the second impression is very often much more correct, the first one is much more bright, because it is coloured by a strong emotional response. The distortions in the perception caused by the first impression as well as the illusion that the Cyberspace is the real life are rooted in the emotional sphere of the person.

The balance between the intellectual sphere and emotional sphere of the person is destroyed, and it leads to either overestimation of the problem or underestimation of possible emotional or intellectual reaction of a person to a challenge of any kind.

Let us consider impressionism under the frame of cognitonics; more exactly, the way it helps to turn a visual feast into a splendid opportunity to understand how strong the first impression is:

- how it provides an opportunity to make children and teenagers see the possibility to view the world in a different way;
- to notice the transient appearances of one and the same object (communicative situation) depending on the season, weather, the background of the viewer, his/her mood, etc.

9 "Secrets" of the Impressionists

One of the "secrets" of the impressionists is hidden in their manner of painting. To enrich the colour of their canvases, the impressionists made use of what is known as division of colour and optical blending. For example, to represent a green meadow, they put little dabs of blue and yellow on the canvas which were supposed to combine to form green in the eye of the spectator.

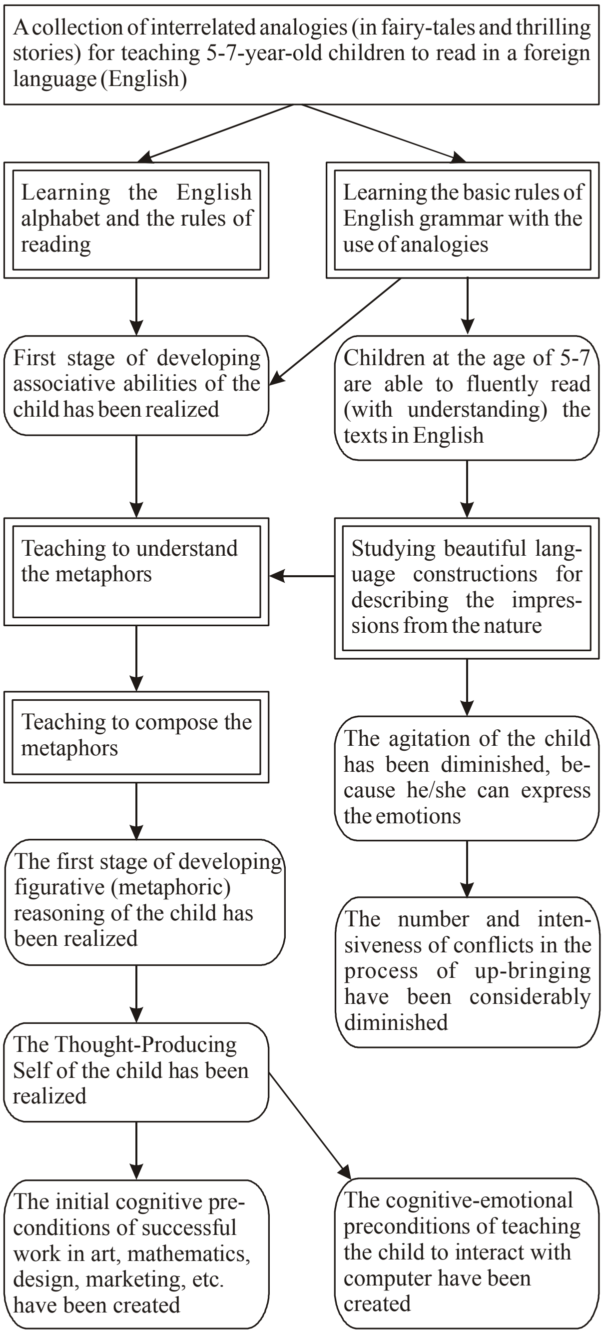

Figure 1: A map of cognitive transformations corresponding to the basic stage of developing creativity.

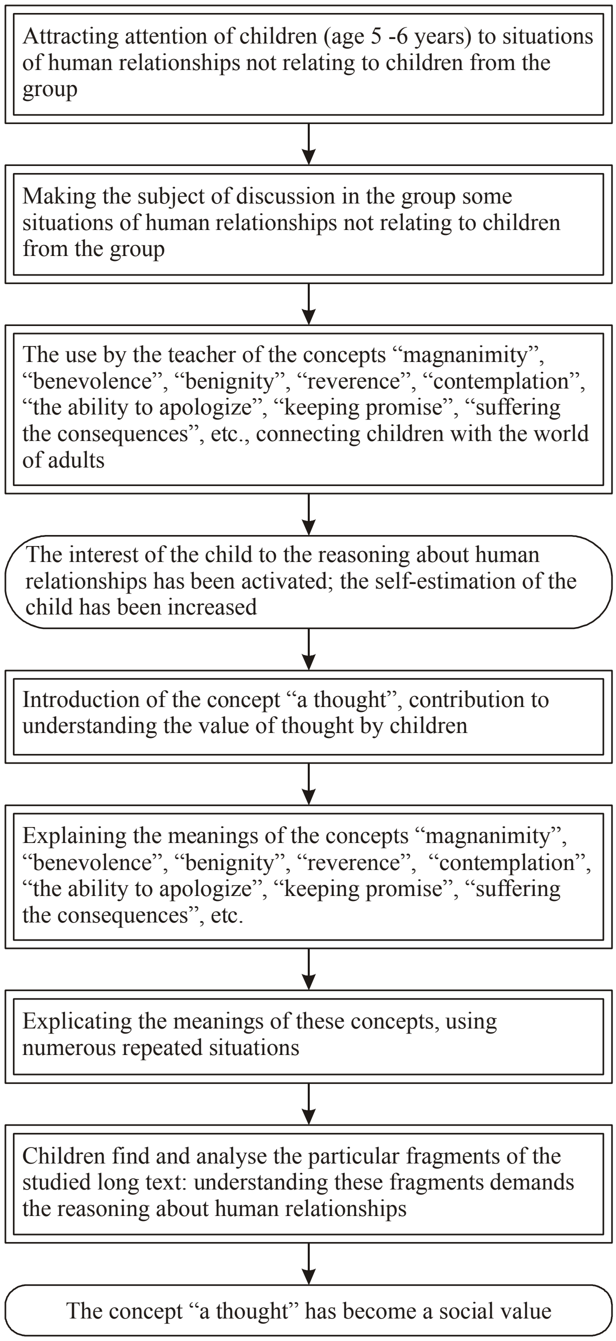

Figure 2: First year of forming a cognitive-emotional subspace of moral values and social responsibility.

The impressionists, for example, Claude Monet, devoted themselves to capturing in paint the fugitive effects of light falling on objects and the play of reflections. Due to this style of painting, the impressionists manage to render the effects of sunlight, vibrations of water and air, the thousand and one reflections on water, etc.

Studying impressionism is a great pleasure. The emotional response of children can’t be overestimated. In case children are asked to come very closely to the pictures, they become deeply impressed by the fact that they can’t see anything except for the mixtures of dabs, a kind of colourful chaos. Watching the pictures from some distance, children come to understanding the beauty of canvases, in fact, they make their discovery of transfiguration. An apparent transformation from the colourful chaos to the visual feast produces a deep lasting effect on the mind and feelings of the child. It is a discovery of an illusion (first impression which is false, though bright).

Take “Water lilies. Green harmony” by C. Monet. In fact, there are no white lilies there, though the water surface is dotted with the common white water lilies.

Children suppose that it is one more example of illusion: we are sure that water lilies are white, but in fact it is impossible to point at and list one white lily. They come to understanding the effect of transparency of the water, on the one hand, and the reflections of the sky, clouds, plants edging the pond, and the whipping willows, on the other hand.

10 Unexpected Conclusions Derived from Studying Impressionism

The acquired experience of perception can be applied to various communicative situations at school with teacher, classmates or any kind of misunderstanding.

When a child or a teenager is at sea or in a fix, he/she is sure that everything is "black". But the idea that he or she hasn't see pure white lilies on the canvas makes him/her think that it may be an illusion, and everything is not so bad, and it is necessary to step aside and look at this situation from some distance.

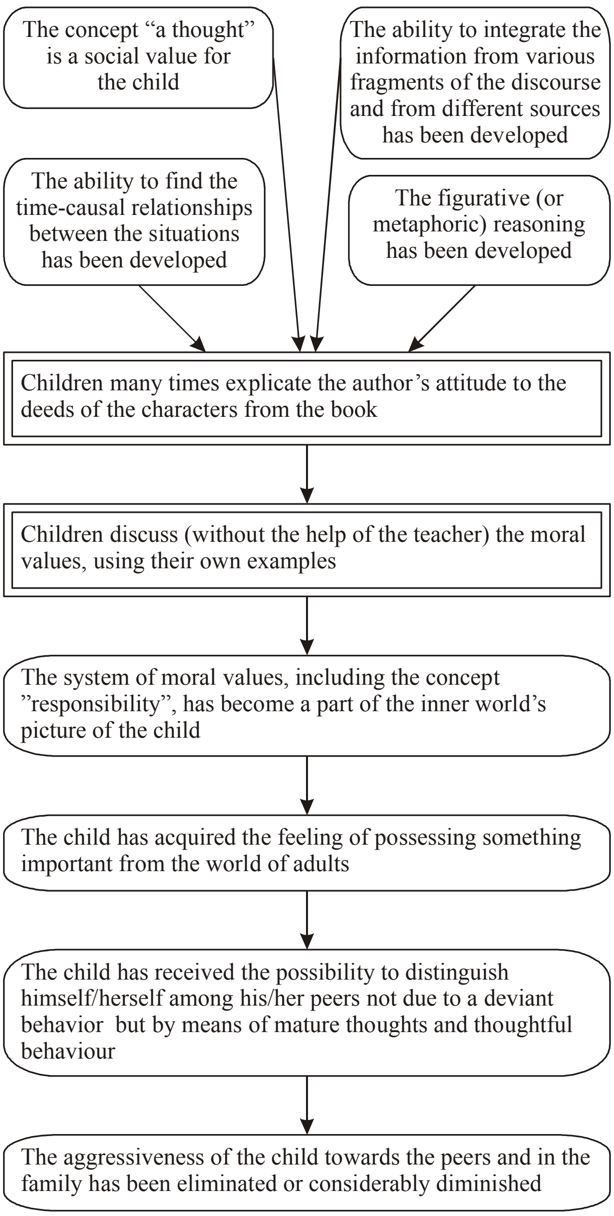

Figure 3: Second year of forming a cognitive-emotional subspace of moral values and social responsibility.

It is the process of establishing a link between the constructed mental representation of the seen pictures and the constructed inner visual image of the life situation. The suggested conclusion: no panic, no chaos, no black situations. It is not passive refection, it is an active transfiguration that makes the life brighter, stimulates a creative response to it.

Another important discovery done by children is connected with the possibility to view the world in a different way and to notice the transient appearance of one and the same object depending on various things.

The ability to see ordinary things in a new way stimulates curiosity and desire to reveal personal perception. It reveals the ability of the child to view the world actively, creating his/her own images, metaphors, corresponding to his/her conceptual picture of the world.

For instance, children are asked to describe the lilies in the pond or the pond itself. The following descriptions were given by children.

Example 1 (a girl, 6 years old): "Near the castle, there was an old park with wide spreading trees and a pond with white lilies. At dusk, lilies are falling asleep, and it seemed that someone had eaten whipped cream from the blue cap of the pond".

Example 2 (a boy, 7 years old): "At night the pond looked like a mirror reflecting bright sparkling stars, and it seemed that one could scoop out a bucket full of stars".

Example 3 (a girl, 7 years old): "At dusk, the pond is fringed with silvery light and looks like an ancient looking glass of the moon being lost by carefree crescent".

The images of the pond are rooted in the own life experience of eating whipped cream or noticing the reflections of the starlet sky in bucket, barrel, or well in the country or playing with a grandmother's looking glass in the silver frame.

Such kind of work proves the importance of the impression as a bright flash evoking emotions, making clear various links between different domains.

11 How to Make Feelings Become the Subject of the Thought

Though impression is formed without conscious thought or specific knowledge, the child should be taught to analyse his/her impressions, to appreciate them. Impression is a kind of impulse sent by the outer world and accepted by the child unconsciously. Impressionism aimed at revealing impression. It provides the splendid opportunity to see the world with the eyes of the painters who were deeply impressed by it and elaborated a special language to express the admiration or just a way of viewing. Such kind of approach to studying impressionism leads to analysing the accepted impulse, it leads not only to emotional response but also to intellectual response. Feelings are becoming the subject of the thought. It may prevent children from impulsive decisions caused by the first impression that looks like an emotional attack.

In the information society, the possibility of such unexpected emotional attacks by social nets is increasing. The way impressionism is taught under the frame of cognitonics is one of the keys to the solution to this problem, because of constructing the mental representation of what is called “the illusion of white lilies”.

12 The Cross-Disciplinary Course “Foundations of Secure Living in Information Society”

The information society we live in has its peculiarities, advantages, and disadvantages, as any other society. In order to speak about successful socialization of children and teenagers, it is necessary to make children understand the ways of living, participating in the social networks, communicating via e-mail, taking on-line courses, receiving information, etc. They should know how to avoid negative “digital” situations or overcome them, how to distinguish virtual reality and emotions caused by that virtual reality from the real life and emotions caused by that life.

Children should be aware that in both cases they should be ready to suffer the consequences of their careless behaviour or ignorance. Children should be taught the rules of acting in the digital space, paying special attention to the moral standards, lest they should hurt somebody’s feelings while communicating with the help of informational technologies (IT). They should understand the power of IT and the responsibility of the users. The teachers should find a correspondence between situations taking place in the real life and similar situation from the virtual reality.

For example, if someone reveals aggression in any way, he/she can suffer the same aggressive feedback, can be hurt. In case of the digital space, not a child or a teenager him/herself is hurt and experiences pain, but his/her feelings are hurt and his/her reputation is in danger because of the quickly spread negative information.

The core idea of the course “Foundations of secure living in information society” is the same for real and virtual life: treat others the way they want to be treated, show compassion and consideration, learn from success and failures, appreciate learning opportunities, etc.

But children should be aware of the difference between person-to-person communication and person – digital environment – person communication, because the digital environment has its own power that can enhance the communicative (and any other) situation.

The goal of the course is to contribute to successful socialization of children and teenagers in the information society. The subject of the course is to introduce students into the digital space, paying special attention to ethics, to the rules of interaction and communication.

Students acquire knowledge of what is strongly prohibited and what kind of behaviour is expected. The course shows the clear difference between virtual reality and reality and establishes correspondence between the ways of perception by people of various situations happened in digital space and real life, on the one hand, and the possible consequences caused by that difference.

The idea of this cross-disciplinary course and ethical approach to its development are suggested under the frame of cognitonics.

13 An Algorithm of Transforming Negative Emotions into Positives Ones

Let us consider the possible reactions of a human (including the recommended reactions) to the so called emotional attacks via social networks. An analysis of the kind could be introduced into the program of the course “Foundations of secure living activity in information society”.

Head Module of the Algorithm
”Processing Messages from Social Networks”

begin
If the first impression (a strong one, it can be either true or false) is POSITIVE then Procedure 1
else { the first impression is NEGATIVE}
Procedure 2
end

Description of the Procedure 1
begin
Make conclusion 1: no harm at the moment of getting an impression;
Make conclusion 2: the feelings are not hurt;
Make conclusion 3: The situation is over
End

Description of the Procedure 2
{The condition of calling this procedure: the first impression is NEGATIVE, that means that the impression causes panic, confusion at the moment of getting it}
begin
FIRST AID: a reminder of the white lilies on the canvas by Claude Monet (if the mental representation is strong and clear);
Make conclusion 1: It is a situation of uncertainty, not apparently a bad one;
Make conclusion 2: The situation needs reflecting, reasoning;
Start some kinds of intellectual activity, thinking over the situation and diminishing emotional activity;
A little later make conclusion 3: The situation is getting much more balanced, less harmful;
Make conclusion 4: The situation is turning into an intellectual riddle;
Make conclusion 5: Now the situation causes another kind of emotions, they are based on the feeling of curiosity;
Make conclusion 6: A situation of another kind has emerged, it aims at solving the riddle

{the transformation of the emotions into the positive ones is over}
End

14 Conclusion

Our study started in early 1990s and including both theoretical and practical aspects has shown that the objectives of cognitonics concerning educational practice are realistic. In particular, taking into account the dangerous sides of acquainting socially inmature children (including teenagers) with modern informational technologies, we argued in this paper the necessity of looking from new positions at the problem of early children's socialization in the computer age. Our vast experience has shown that it is possible to do a lot for eliminating or considerably diminishing the aggressiveness of children, using traditional teaching materials (the fairy-tales, thrilling stories, etc.) but in a new way: extracting from these sources the attitude of the author to the deeds of the characters from the text and forming step by step the cognitive subspace of moral values and social responsibility. The essence of our approach to early socialization of the child was explicated with the help of three maps of cognitive transformations (MCT). It seems that the notation of MCT will be of help for the scholars elaborating the methods of contributing to the well-balanced, harmonic development of the personality in information society.

The development of civilization is the endless process of challenges and answers. Internet, new informational technologies are a challenge. It is not only an intellectual challenge but a spiritual challenge as well. The illusion of the true existence of the Cyberspace gives birth to new kinds of emotional attacks via e-mail, social nets, and cell telephones. It is difficult to resist these attacks, because lots of teenagers and grown-ups become aware of the information together with the attacked child.

Impressionism as a manner of painting, rooted in the idea of the first impression, being taught under the frame of cognitonics helps to construct the vivid mental representations of the illusive situations in the minds of children and teenagers. The example considered in this paper is positive, impressive, and it is based on children's life experience (they acquired it while watching the painting). On the other hand, they have already thought about their own examples taken from the real life: one child is sure that the dog is angry, because it barks; another child says that his particular dog is kind, because it wags its tail every time the child sees it.

To make children and teenagers understand how the illusion and first impression work, explaining to them emotional constituent of these notions, providing them with thrilling and clear examples is very important, especially in the information society when they deal not with one partner of communication but with many partners from the nets.

People communicating via social networks don't take into account the child's mood, character, the events of the day, child – parents relationships at the moment, background of the child. New possibilities of ICT demand much more developed ability of the child to resist to any reply or replies, much stronger confidence in oneself, and clear understanding how the illusion works.

We know that the lilies on the pond are white, but, in fact, there are no white lilies in the pond: everything depends on the maturity of the viewer.

15 References

[1] D.D. Burdescu, M.C. Mihaescu, M.C. Ionascu, I. Buligiu, B. Logofatu (2010). Expanding mental outlook with the help of concept maps. *Informatica. An Intern. Journal of Computing and Informatics (Slovenia)*, 34 (4), pp. 535-540.

[2] Dagdilelis, V. and M. Bontila (2011). Lost in Cyberspace(s). *Proceedings of the International Multiconference Information Society - IS 2011, Slovenia, Ljubljana, 10 – 14 October 2011. The Conference Kognitonika/Cognitonics*. Jozef Stefan Institute, 2011, pp. 376-379; available online at *http://is.ijs.si/is/is2011/zborniki.asp?lang=eng*.

[3] V. Fomichov and O. Fomichova (1994). The Theory of Dynamic Conceptual Mappings and its Significance for Education, Cognitive Science, and Artificial Intelligence. *Informatica (Slovenia)*, 18 (2), pp. 131-148.

[4] V. Fomichov and O. Fomichova (1997). An Informational Conception of Developing the Consciousness of the Child. *Informatica (Slovenia),* 21 (3), pp. 371-390.

[5] V. Fomichov and O. Fomichova (1998). A new theoretical and practical approach to early positive developing child's consciousness. In *R. Trappl (Editor), Cybernetics and Systems'98. Proceedings of the 14th European Meeting on Cybernetics and Systems Research. Vol. 1*, Austrian Society for Cybernetic Studies, Vienna, pp. 276-281.

[6] V. Fomichov and O. Fomichova (2000). The social responsibility of computer science specialists for the creative potential of the young generation. *Intern. J. of Artificial Intelligence in Education. Special Issue on AIED 2010*, 11 (2), 208-219.

[7] V. Fomichov and O. Fomichova (2006). Cognitonics as a New Science and Its Significance for Informatics and Information Society. *Special Issue on Developing Creativity and Broad Mental Outlook in the Information Society (Guest Editor Vladimir Fomichov), Informatica. (Slovenia)*, 30 (4), pp. 387-398.

[8] V. Fomichov and O. Fomichova (2011a). Preface/Predgovor. Second International Conference on Cognitonics (Cognit 2011). In *Proceedings of the International Multiconference Information Society – IS 2011, Slovenia, Ljubljana, 10 – 14 October 2011, Vol. A.* Jozef Stefan Institute, 2011, pp. 349-350; available online at *http://is.ijs.si/is/is2011/zborniki.asp?lang=eng*.

[9] V. Fomichov and O. Fomichova (2011b). A Map of Cognitive Transformations Realized for Early Socialization of Children in the Internet Age.

Proceedings of the International Multiconference Information Society - IS 2011, Slovenia, Ljubljana, 10 – 14 October 2011. The Conference Kognitonika/Cognitonics. Jozef Stefan Institute, 2011, pp. 353-357; available online at *http://is.ijs.si/is/is2011/zborniki.asp?lang=eng*.

[10] O. Fomichova (2009). *Humanitarian Education – an Answer to the Challenge of Time*. Moscow State University Press, Moscow – 368 pp. In Russian.

[11] O. Fomichova and V. Fomichov (1996). Theoretical foundations of a new method of teaching children effective information processing. *Informatica (Slovenia)*, 20(3), pp.381-399.

[12] O. Fomichova and V. Fomichov (2000). Computers and the Thought-Producing Self of the Young Child. *The British Journal of Educational Technology*, 31 (3), pp. 213-220.

[13] O. Fomichova and V. Fomichov (2007). Cognitonics as a New Science and Its Social Significance In the Age of Computers and Globalization. *IIAS-Transactions on Systems Research and Cybernetics*. Vol. VII, No. 2. Intern. Journal of The International Institute for Advanced Studies in Systems Research and Cybernetics. Published by The Intern. Institute for Advanced Studies in Systems Research and Cybernetics (IIAS), Tecumseh, Ontario, Canada, pp. 13-21.

[14] O. Fomichova and V. Fomichov (2009). Cognitonics as an Answer to the Challenge of Time. *Proceedings of the International Multiconference Information Society- IS 2009, Slovenia, Ljubljana, 12 – 16 October 2009. The Conference Kognitonika/Cognitonics*. Jozef Stefan Institute, 2009, pp. 431-434; available online at *http://is.ijs.si/is/is2009/zborniki.asp?lang=eng*.

[15] O. Fomichova and V. Fomichov (2011). Impressionism in the Mirror of Cognitonics. *Proceedings of the International Multiconference Information Society - IS 2011, Slovenia, Ljubljana, 10 – 14 October 2011. The Conference Kognitonika/Cognitonics*. Jozef Stefan Institute, 2011, pp. 380-383; available online at *http://is.ijs.si/is/is2011/zborniki.asp?lang=eng*.

[16] J.D. Novak (1998). *Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations*. Lawrence Erlbaum Ass., Mahwah, NJ.

[17] H. Yasuda (2009). An Information System in School for Risk Management of the Internet: Preventing Cyberbullying Without Prohibitions. *Proceedings of the International Multiconference Information Society – IS 2009, Slovenia, Ljubljana, 12 – 16 October 2009. The Conference Kognitonika/Cognitonics*. Jozef Stefan Institute, 2009, pp. 435-439; available online at *http://is.ijs.si/is/is2009/zborniki.asp?lang=eng*.

Information Technology for Management and Promotion of Sustainable Cultural Tourism

Marija Valčić and Lana Domšić
College of Business and Management "B.A. Krčelić"
V. Novaka 23, 10240 Zaprešić, Croatia
E-mail: marija.valcic@vspu.hr, lana.domsic@vspu.hr

Keywords: sustainable tourism, cultural heritage destinations, e-Tourism, ICT, destination management systems

Received: December 20, 2011

This paper suggests that many of the negative effects of globalization and inadequate tourism growth can be compensate by the use of intelligent ICT solutions in the development of heritage tourism. The encounter between cultural tourism and information and communication technologies represents an opportunity to preserve national culture, create partnership and enhance destinations value in information society. To this aim, we propose a model for local Destination Management Systems of heritage destinations (DMS) which could help to achieve a more globally responsible paradigm for the tourism industry and facilitate the management of destinations and the coordination of the local suppliers. DMSs provide interactive demonstrations of local amenities and attractions and enable consumers to build their own itinerary based on their interests and requirements. All the stakeholders within the destination are linked with each other in order to create collaborative action and a genuine, sustained growth in heritage tourism. An example is given of the Croatian World Heritages Sites and their status on the Web.

Povzetek: Predstavljene so informacijske tehnologije za podporo trajnostnemu turizmu.

1 Introduction

The globalization and changes in modern society represent both opportunity and threat for national cultures, heritage and identity. Cities and regions are losing their traditional heritage and keeping up with global trends and fashions rather than reviving local traditions, history and values. By doing so they become less attractive places to live, work, visit, and invest in. Therefore, economic development driven by international tourism, especially in smaller communities, may be short lived and have many negative effects.

The tourism industry, facing its everlasting dilemma of growth versus sustainability, can be regarded as contributing to critical trends in world development rather than an instrument of world peace and tolerance (as praised by international organizations like UNESCO). It is clear that the traditional approach pursued by the tourism business is in need of fundamental revision. The increased cultural diversity of today's world, together with the growing access to the heritage through the Internet, is likely to result in strong pressure for its fundamental restructuring. A technological link should be established between heritage and a sustainable tourism economy based on the cultural richness of places. Heritage should be considered as a bridge between the past and the future of a community, a reflection of founding values, history, and identity [3]. The cultural capital embodied in buildings, artefacts, sights, songs, and rites permits the transmission of the culture of a people through time and space. The heritage industry, small in size, information-intensive and creative provides a means to interact and to learn about host communities.

ICTs and Internet can favour the reconciliation of heritage and tourism, supporting a process of the empowerment of local stakeholders and of creative encounter between host and guest communities. This progress should support creativity, collaboration and appreciation of national and world heritage, taking into consideration the cultural and social capital that is necessary for sustainable community economic development.

2 Globalization and its Impacts on Tourism

Globalization has profound implications on competitiveness, trade and tourism policy. In an increasingly global and competitive market characterized by standardization, cultural crisis and pollution, tourism is today more than ever regarded as a "problem area," something to constrain and regulate, rather than a strategy to pursue cross-cultural integration across the world [3]. Technical progress in transportation has enhanced the physical accessibility of destinations, but this has not been matched by an equal increase in cultural access, the subjective capacity to recognize and attribute a value to the cultural features of the visited places.

There has been shift toward commercialization and short-term decision making at the expense of conserving cultural heritage and, most often, tourism development has resulted in increased strain from tourist pressure on host communities. It fails to contribute to the elevation of their economic status and shows to be a short-lived option for development. There are many negative effects of tourism for the socio-cultural environment of community: loss of identity and culture wich in many cases takes form of westernisation; modification of traditional art forms and culture (building, styles, systems, clothing, events); commodification of performances and ceremonies in order to appeal to tourists; environmental concerns, pollution and loss of natural and architectural heritage; seasonal and unstable jobs for locals.

For the area of heritage tourism, a new model is necessary that joins contributions from local institutions and private enterprises that are looking for ways to build on their existing potential, organizing them developing and enhancing local cultural assets, emphasizing the unique character of the place [3].

Information and communication technologies are developing and expanding extremely quickly and have a huge impact on tourism. Rapid advances in ICTs challenge the tourism and travel industry at many levels because they deeply affect the organization and governance of tourism and travel value chains and thus the economics of the industry [4]. The transition to an information society creates new opportunities and threats and offers individuals a range of choices to be eclectic. Both information technology experts and social scientists should show interest in these issues, because the long-term success of their field of inquiry depends largely on its ability to bring about integration at the host community level, enabling the public and private sectors to cooperate and use local resources for development efficiently.

3 ICT Solutions in Achieving Sustainable Heritage Tourism

A key challenge for the enhancement of cultural heritage is to bridge the different perspectives of how tourism organizations and service providers can present their cultural heritage in a way that appeals to the interests of the international tourism audience. The goal is to generate value from local knowledge and information and make it available to consumers worldwide: virtually, via the Internet, and direct physical access. The use of ICT is necessary and it involves various stages of the operation of the heritage industry like content creation and communication, value enhancement, and market strategy. A basic prerequisite for sustainable tourism is allowing individuals and communities an opportunity to be included and connected. There is a need to develop the aspects concerning the use and development of tools, technologies, and methodologies to facilitate the efficient networking of information and communication systems in tourism. Utilizing the ICT infrastructure, communities are helped to become more strategic and entrepreneurial in managing their heritage [3].

New technologies can produce an essential contribution to tourism development. Cultural heritage tourism is increasingly depending on ICTs for purposes of promotion, distribution and delivery of products and services. It also provides a tool for communication between tourism suppliers, intermediaries and consumers. Web-based visitation is becoming commonplace as the tourism industry adopts networked interactive multimedia technologies. The adoption of Web technologies is affecting the ways that tourists become aware of destinations, the ways they select and experience destinations.[5] Multimedia is becoming one of key areas of development that influences tourism. Tourism information needs an extensive representation of photos and graphics in order to provide a tangible image or experience to travel planners. Using animations or video clips can enhance information richness and interaction [6].

E-Tourism is maturing fast as a mainstream distribution mechanism, therefore establishing Internet presence and e-commerce strategies will become critical for destinations to remain competitive. It is clear that cultural heritage is not well represented in existing e-destination platforms and e-tourism services. Establishing close links between the cultural heritage and ICTs is crucial to innovation in heritage destination management. Electronic reservation systems used in today's tourism industry are being transformed into integrated Destination Management System (DMS).

DMS is a set of available interactive digital information about the destination and its related products and services which provide "total tourism product" or "travel experience" [7].

It gives users an access to a comprehensive picture of a destination touristic product, through different channels and platforms, providing extensive information on destinations and attractions, as well as the ability to perform search and booking in real time. DMS also serves as a destination management tool, a tool for marketing and promotion, and support for small and independent providers of tourist services [8].

Furthermore, technological developments can be expected to lead to multichannel multimedia DMS

Figure 1: Destination Management System.

serving purposes not only of travel information distribution, planning, and fulfilment, but also of travel-related education and entertainment.

ICT also enables heritage sites to expand their activities in the geographical, marketing and operational sense and play a particularly important role in managing relationship with customers [9]. Much information about cultural heritage already exists in publicly available resources, but can also be incorporated into more structured promotional and educational programs and platforms. The objective is to provide e-content for travel planning and education. Technical partners need to be brought together with local tourism organizations to design and implement Internet sites and portals that will aggregate information and services related to the cultures of each host city in terms of historical perspective, architecture, landscape, fashion, cuisine, and public culture. This joint-venturing process needs to be actively stimulated and supported by the local and regional governments and other nongovernmental organizations, in particular in regions that are now "disconnected." A further feature would be to elicit comments from tourism clients that navigate these online materials as to how they could be improved. In this way, guidelines for evolutionary improvement can be obtained [3].

Web page of the site can be used in various ways and for different purposes including preservation, education and site management. It educates visitors about the site and need for its protection. It enables market segmentation and gives a costumer relevant information, contextualized and supplemented with history facts, stories and related objects and sites. Hypertext links can channel the different parts of the audience to different places on the page (for example, parts specifically designed for children. The sites often have limited space for exhibition and must choose the artefacts that will be exposed while the rest remain in storage. Unlike the real-space sites, web site virtual space is infinite and can be used to display objects that are not exposed at the site. Virtual visit allow access to the site for those audiences who have no possibility to travel to the destination and thus the sites better meet its mission of enabling public access. A unique attribute of heritage consumption is that its benefits are experiential and may be divorced from the place itself [10].Virtual travel experiences are especially appropriate in the case of heritage sites in which physical visitation is discouraged in order to conserve the resource or is not possible for financial or other reasons.

Furthermore, Web sites play an important role in the creation of sustainable tourism. An essential element of sustainable development of cultural tourism is the behavior of visitors at the site. There are different channels for raising awareness. Quality of information about heritage encourages visitors to understand the characteristics of heritage and the need for its protection and helping visitors to enjoy the site appropriately. On the other hand, the promotion can significantly help in achieving financial and educational goals. Good presentation on the Internet can attract more visitors, if the site carrying capacity allows it. Then it increases the profit that can be used to fund educational activities, solve management problems and reach the goals and objectives of the site.

ITC solutions are not necessarily expensive for a heritage destination. Joining with other destinations in a common web page can significantly reduce costs and also better respond to customer needs. Typical heritage destination DMS should include interactive maps, 3D applications, virtual tours, online exhibitions, interactive learning resources, games and fun tools, online collections and databases, user communication, community aspects, personalization and online shops. It can be developed in the collaboration with educators, art historians, historians, artists, museum specialists, etc. It should also provide accommodation reservation and events bookings, personalized navigation, interactive maps, travel journey planner about local weather and public transport.

These identified e-Services can vary in their technological development, from some simple interactive structures to more complex interconnected e-services, not only using different media, but also linking and

WEB SITE FUNCTION	CONSERVATION	EDUCATION	DESTINATION MANAGEMENT
DESTINATION AWARENESS	Increases awarness to reduce negative impact	Supplements knowledge before, during and after the visit	Reduces staff's time answering the publics' questions
INFORMATION PROVISION	Opportunty to present conservasion message	Visitors can access the information they choose according to market segments	Site managers restrict access to fragile areas
INVENTORY AWARENESS	Using digital image instead of fragile artefacts	Shawcase entire inventory and relate with relevant context, stories, sites	Addresses accessibility issues and provides better capacity management
VIRTUAL TOURS	Restrict public from fragile areas	Provide edutainment, entertainment and education combined	Ensures every visitor sees the same reconstuction

Table 1: Heritage Web Site function (adapted from Buhalis, 2008).

displaying the different e-services together. These portals perform two functions. Firstly, they can be seen as a marketing tool that projects the region in perspective to the rest of the world. Secondly, by means of educating tourists about the prevailing culture of the region, a more satisfactory travel experience can be offered.

The value enhancement process is expected to produce a number of outcomes, which can be categorized as "guest satisfaction," "profitability," and "sustainability." These three categories result from the encounter of tourism demand with the supply of local culture. The centre of attention is still the tourist; yet he or she is no longer a passive participant in preconceived itineraries, but rather an aware, active player in the process of codetermination and reward of cross-cultural experiences. Guest satisfaction results from the intensity and quality of the cultural experience [3]. Satisfaction is also clearly connected with profitability, through the mediation of perception and commercial strategies—hence the importance of cooperation at different levels in destination marketing.

This kind of development should lead to more sustainable and responsible tourism that will be able to maximise economic benefit to local people and enhance their quality of life and at the same time build local pride and confidence. It will develop quality tourism products by supporting small and micro industries and involving local people in decision-making. Finally it will lead to promotion of understanding and better interaction between local people and tourists and promotion of ethical values common to humanity.

The main goals are to achieve eco-tourism on the basis of carrying capacity of the sites, respect to artistic, archeological and cultural heritage. Financial resources from tourism industry should be used for maintenance of sites and monuments, encouraging survival of traditional cultural products, crafts and folklore

4 Example of World Heritage Sites in Croatia

For the purposes of this paper we conducted a research of the Web presence of Croatian cultural heritage tourism based on UNESCO World Heritage sites, showing the heterogeneity of this segment of the travel industry and the diverse origins and forms of its presence on the Internet. The analysis showed that most of the destinations are largely invisible on the Web and few provide web-based access to travel planning services. The visibility of a World Heritage Sites on the Internet is key factor in their emergence as a virtual heritage destination that influences the actual physical visitation. The cultural heritage tourism sector in Croatia seems slow to adopt new technologies. There are a number of barriers such as the low level of cooperation between stakeholders, the lack of the strategic vision and business planning and the limited levels of understanding of the eTourism potential. This is affecting the presentation of the country on-line as there is currently no Destination Management System or a comprehensive portal that promotes cultural heritage destinations in Croatia.

World Heritage Sites (WHS) are considered to be the centrepiece of the global heritage tourism industry. The original purpose of designation as a WHS under UNESCO's Convention was to assist with management and preservation of the cultural heritage site and to encourage the development of management plans. However, it is believed to increase tourist visitation [11] and many WHS are becoming major cultural tourism attractions. The inclusion of heritage sites in the World Heritage List provides a powerful branding that helps to preserve and promote the destination and facilitate the development of their resources for tourism [12].

First of all, we wanted to see who puts the information about the Croatian World Heritage sites on the Internet, that is, who are participants in creating images of these heritage destinations. We used Google, currently the most common search engine on the Web, and the term "World Heritage Sites in Croatia" in both English and Croatian. We analysed the data collected from the first hundred pages that have appeared in Croatian and in English and we classified these pages in several categories of heritage tourism stakeholders. The result of the analysis is shown in percentages in the table below.

The analysis shows that the intermediaries have a predominate role in the dissemination of information on world heritage destinations in Croatia. Individual web sites offering information on WHS are set up by a wide variety of organizations with overlapping jurisdictions, including tour operators, national, regional and local organizations, and of course the operating agencies of the sites themselves. These actors are promoting destinations for different reasons, but their shared interest is to increase awareness of the WHS, although not necessarily to maximize physical visitation.

There are a major number of national tourism organizations web pages which are often devoted to their

Web Page	Search results In Croatian	Search results In English
Tour operators and agencies	16%	17%
Regional and local destinations	12%	2%
National portals, tourism associations and organizations	23%	15%
Media and publications	23%	9%
Events and conferences	2%	0%
Academic and educational sites (Universities, libraries, encyclopedias)	4%	10%
Social web (blogs, wikis, social networks)	4%	17%
Commercial service providers (hotels, restaurants, renting)	14%	10%
Individual attractions (sites, museums, monuments)	2%	0%
International tourism portals and on-line guides	0%	20%

Table 2: Composition of Heritage Tourism Web pages.

development plans, programs, and projects. There are also a large number of web pages of tourism and cultural web portals and various media. In second place we found, in almost equal proportion, local and regional tourist offices, tourist agencies and service providers that help promote the destination with the main goal to increase the number of visitors.

The biggest promoters of our heritage sites in English are different international portals and guides. A significantly large role has a social web where users themselves public information via blogs and social networks. An important fact is that only 2% of the web pages in Croatian and 0% of pages in English are dedicated to individual heritage sites, which means that the majority of destinations don’t have their own websites. Also there is no major common web site devoted exclusively to World Heritage Sites in Croatia. The only two destinations that have their own web sites are Plitvice Lakes National Park and the Stari Grad Plain in Hvar. These two pages contain options in several languages, reservations systems, information on local weather and pictures or videos of the site. But they still lack a number of elements to become the real DMS.

To fulfill their function, these web pages will have to include elements such as internal search engines, information that allows easy organization of travel, references to local carriers, hotels, restaurants, activities and agendas of local events. Multilingual promotion is also necessary to position a site in specific travel markets.

The quality destination management system will depend, of course, on the maturity of a particular locality, the general awareness among the passengers of the heritage location brand and the degree of development of

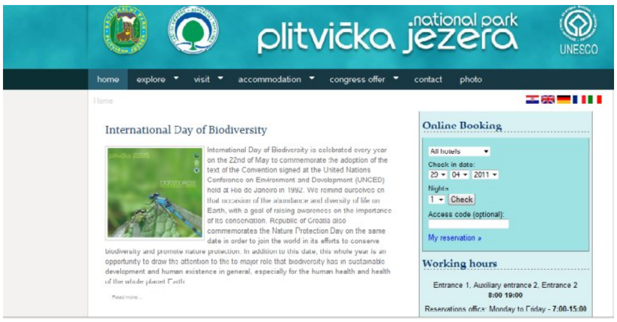

Figure 2: Web page of Plitvice National Park, Croatia.

Figure 3: Web page of Stari Grad Plain in Hvar, Croatia.

heritage tourism industry and tourism products and services around the site. For example, Dubrovnik as a historic city already possesses a well developed tourism infrastructure, specialized service providers and additional cultural attractions. Only the technology infrastructure lacks to form a DMS. It is important that persons responsible for site management embrace ICT as an important and necessary element in fulfilling the tasks of protecting, the public presentation and communication of heritage sites.

5 Conclusion

Cultural tourism development currently presents some very definite unbalances. On one side, it depends on localized and hardly reproducible resources. On the other, it is governed by an industry that is increasingly both global in nature and disconnected from the sources of cultural capital. Like any industry, tourism needs profit and investment incentives to grow, but both commercial interests and government entities should work to achieve a reconciliation of the inherent conflict between heritage and tourism. The potential from tourist growth can only be fully exploited if both policy makers and businesses remove unnecessary structural barriers to growth, by capitalizing on the opportunities that are based on cultural heritage and identity [3]. In other words, a new strategy is required for developing cultural heritage as a viable economic sector. With time, there should be a change from the administered industrial economy to an entrepreneurial economy accompanying and institutionalizing the information society.

Heritage tourism is currently in the process of systematization of information, communication and multimedia as the means to achieve a competitive advantage and sustainability. Awareness and knowledge about a destination and perceptions about its quality and value are factors that influence the motivation of visitors and the selection of a destination. Destinations will gradually have to incorporate the online Destination Management System (DMS) that can improve promotion and management using integrated e-services. New Destination Management Systems can convey diverse, comprehensive and multimedia information on heritage destinations and surrounding products and services, and thus contribute to the destinations long-term competitiveness and sustainability.

In Croatia there are no DMSs at the national level or for the particular destinations, and cultural tourism in Croatia has not yet started to widely use information, communication and multimedia technologies. UNESCO sites are the main points of heritage tourism in Croatia. They don’t have their own web pages, but are present at other web sites of special interests that provide limited information. The best solution would be a portal that would unite all destinations in one place and offer a complete information and experience as a base of heritage tourism. Collaborative networking should be established between all stakeholders and both the policy makers and the entrepreneurs should work together to raise awareness of e-tourism through training and education.

References

[1] UNESCO. (2001). Message from the Director-General of UNESCO, Mr. K. Matsuura, at the Official Celebrations of World Tourism Day, September 27, 2001, on the theme: “Tourism, a tool for peace and dialogue among civilizations”[On-line document]; http://www.unesco.org/ dialogue2001/en/dgtourism.htm

[2] Graham, B., Ashworth, G. J., and Tunbridge, J. E. (1998). A Geography of Heritage: Power, Culture and Economy. London.

[3] Go, F.M, Lee, R.M., and Russo, A.P. (2003). e-Heritage in the Globalizing Society: Enabling Cross-Cultural Engagement Through ICT. J. of IT & Tourism, 6(1), pp. 55-68.

[4] Buhalis, D. (1998). Strategic Use of Information Technologies in the Tourism Industry, Tourism Management, 19(5), pp. 409-421.

[5] Davis, A.,Prentice, R. (1995). Conceptualizing the Latent Visitor to Heritage Attractions, Tourism Management, 16(7), pp. 491 – 500.

[6] Buhalis, D. (2003). eTourism: Information Technology for Strategic Tourism Management, Pearson (Financial Times/Prentice Hall).

[7] Buhalis, D., Spada A. (2000). Destination Management Systems: Criteria for Success – an Exploratory Study. Information Technology and Tourism, 3, pp. 41-58.

[8] Frew, A.J. and Horan, P (2007). Destination Website Effectiveness – A Delphi Study-based on eMetric Approach, Proceedings of the Hospitality Information Technology Association Conference, HITA 07, Orlando, USA.

[9] Buhalis, D., Owen, R., Pletinckx, R. (2006). ICT applications for World Heritage Site Management, in Leask, A., & Fyall, A. Heritage Site Management, Butterworth-Heinemann, London, pp.125-144.

[10] Prentice, R., Duncan, L. (1994). Market-based Product Development in Heritage Tourism. Tourism Management, 15(1), pp. 27-36.

[11] Shackley, M. (1998). Visitor Management: Cases from World Heritage Sites. Oxford: Butterworth Heinmann.

[12] Drost, A. (1996). Developing Sustainable Tourism for World Heritage Sites. Annals of Tourism Research, 23(2), pp. 479-492.

Linguistic Model Propositions for Poetry Retrieval in Web Search

Magdalena Granos and Aleksander Zgrzywa
Wroclaw University of Technology
Wybrzeże Wyspianskiego 27, 50-370 Wrocław, Poland
E-mail: magdalena.granos@pwr.wroc.pl

Keywords: linguistic models, poetry retrieval, information retrieval, web search

Received: January 23, 2012

The paper investigates the linguistic models for poetry retrieval purposes and raises the importance of access to advanced exploratory search engines, which, according to the Maslow's hierarchical pyramid of needs, address the needs of higher order associated with self-fulfilment and satisfaction of informational-cognitive aesthetic ambitions of non-verbal transmission. This task statement is formulated in the context of a new scientific discipline – cognitonics. The work focuses on filling the information needs of sublime poetry as a form of structurally categorized pattern, which is an important indicator to determine the search methods of poetry based on adaptation of linguistic models. The issues addressed refer to probabilistic models that allow for predicting the occurrence of words in a sentence on the basis of distance functions relatively to similarity of words and phrases, as well as by k-nearest neighbours strategy, and frequency of words relatively to rankings. Moreover, the paper aims at showing the optimal performance of linguistic models in search of the most effective methods of poetry retrieval. During these searches, the relevant information retrieval models reveal their role in terms of new opportunities for Internet search engines into the formative process of education and creation. Accordingly, the paper attempts to explore ways of increasing web-search efficiency to make future research flexible, yet precise, in interpretation of queries.

Povzetek: Opisana je gradnja in uporaba lingvističnih modelov za iskanje pesmi na spletu.

1 Introduction

Amid escalating resources of complex data in website network, the adaptation of search tools for information from the extensive resources of knowledge, implies the inevitability of working towards increasing power search systems. Globalization process of transformation in socio-economic context, calls for smooth availability and flowing information, as well as an operative technology to provide efficient multi-contextual information services for efficiently informed society. Widely and easily accessible information is a common and inexhaustible wealth of human imagination and intellectual achievements, which are inalienable in consciousness in comparison to the resources of derivatives of life.

The information requirement is the space of human knowledge, experienced consciously as a query, leading to look for relevant reply. As far as a user of information system is considered, a measure of meeting the need for information involves the assessment of user knowledge about information systems and services. It also results from the size of a subset of user's population, who positively evaluate a system. Eventually it concerns the extent to which people use different types of services, resources and information collections [17].

During last decades, especial after the birth of the Internet, people in developed countries have been exposed to permanently increasing informational and time pressure, have been forced to take much more decisions in the same time as they used to do before. As a result, people have now much less time for the spiritual-emotional development, and it is one of negative, broadly observed shifts in modern information society. That is why a very significant problem for the humanities, computer science, and web science is to find new ways of supporting spiritual-emotional development of the personality in the information society.

The analysis of the mentioned and some other negative shifts caused by the progress of information and communication technologies and globalization processes, the desire to make something constructive for compensating the discovered negative shifts have caused the birth of a new scientific discipline called Cognitonics [8, 9].

It should be underlined in the context of this paper that Cognitonics formulates a new, large-scale goal for the software industry and Web science: "to develop a new generation of culture-oriented computer programs and online courses (in the collaboration with educators, linguists, art historians, psychologists) - the computer programs and online courses intended for supporting and developing positively-oriented creativity, cognitive-emotional sphere, the appreciation of the roots of the national cultures, the awareness of the integrity of the cultural space in the information society, and for supporting and developing symbolic information processing and linguistic skills, associative and reasoning abilities of children and university students [8].

Along with the growing expectations of information, there is a pressing need for advanced exploration capabilities of Internet search engines, which take into account not only the wide spectrum of phrases, determined by semantics terms from surrounding reality in context of accompanying circumstances. According to the hierarchical pyramid of needs by Maslow as a consequence of once fulfilled expectations, owing from the vital functions, aspirations of a higher order arise, associated with self-fulfilment and satisfaction of informational-cognitive aesthetic ambitions [14]. Poetry, pervading the most subtle areas of human sensitivity, fills the most sophisticated information needs, and becomes an area of particular desirability to find information about non-verbal lore.

2 Concept of Information Need

The concept of information need concerns the gap in a person's knowledge. The information request is expressed as a query asked for relevant sources. An inquirer obtains an answer from an information system.

The substance of an information need is determined during an information interview, giving the recommendations of relevant sources.

In case of a user of information system, a measure of meeting the need for information is associated with:

- an assessment of user knowledge about information systems and services;
- the size of a subset of the population of users, positively evaluating a system or its components;
- the extent, what people use different types of services, sources and collections of information.

We live in the Age of Information and rapid communication. Intensifying the size and flow of information, determines more sophisticated expectations.

The more accessible information becomes, the more possessive and irresistible become expectations for ubiquitous and incessant access in a variety of situations [13] in consequence of claim for advanced culturally, fanciful and subtle areas of information with ambiguous contexts.

Following the Maslow's pyramidical hierarchy of needs, directly from the bottom of basic physiological needs, successively across safety, belonging, esteem, up to the peak of the self-actualization, meant by creativity, intuition, identity and purpose, once the basic needs are met, information becomes more valuable in personal growth.

Such tendencies have continuously claimed research and development of the information and search systems, to fulfil the desire for the most advanced and sophisticated culturally spaces of information, such as fanciful and subtle field of poetry based on nonverbal needs of a user.

Mehrabian's communication model, where only 7% are the words of verbal communication, while 38% and 55%, respectively, constitute intonation and expression of non-verbal communication, shows the validity of new trends development of search mechanisms [16]. They have intuitively interpreted the meaning of messages, which are expressed in poetry through stylistic means, such as onomatopoeic words, accent, rhythm, rhymes, expression of feelings, mood, metaphor, the spatial organization of words and thematic motifs as well [3].

Although being non-verbal in communication, poetry, like musical notation, also appears as a structured form, that is operating according to the categorized models, such as the rhythmic pattern, metric foot, rhyming pattern, versification pattern, alliteration, which are effectively used to develop a search method, submitted in consequence of building a new linguistic model based on structural patterns of poetic works, and by adapting some linguistic models [7].

The process of searching for poems in web browsers, concerning the semantic and syntactic patterns of poetry means, makes the essence of the poetry retrieval.

3 A Philosophy of Language Approach

Philosophy of language has generally been concerned with language and meaning on focus of empirical investigation in linguistics as scientific study of natural language. This is generally referred in philosophy to as the ‘Linguistic Turn’.

Reviewing the philosophic configurations of the language in the twentieth century in the context of life-meaning aspects of special significance are the Chomskian linguistics, which are particularly relevant in a perspective derived from the Cosmonomic Philosophy of Dooyeweerd. Noam Chomsky perceives language as a supervening property of a certain system found in human brain - he calls this a Language System - which development's determines individual and its own internal language as a principal object of study [5].

Herman Dooyeweerd extends his idea of Multiple Rationalities through a new philosophic approach to such issues as meaning, knowledge, being, time, functioning, normativity, theory, practice or social structures, providing grounds for treating life-meaning and the life-world [6], where all human activity is multi-aspectual. Dooyeweerd believes that the kernel meaning of various aspects can be grasped by our own intuition. For example, a concept of justice is imprecise and often divisive but it can also be understood very intuitively.

Similarly, reflecting on aspects of natural laws governing Cosmos, one can intuitively extend certain imperatives into various concepts, faiths or languages. Thus, reflecting on aspects of poetic language, aside of often intuitive and unquantifiable beauty, we can also asses a poem in normative forms, since a poem functions formatively (it has a structure), lingually (it is writing), aesthetically (harmony of style and play of words), juridically (copyright), and so on. Certain aspects are more important than others: the poem can still be a moving expression of aesthetic writing, even without copyright.

The aesthetic and lingual aspects seem particularly important for a poem, though in different ways. The lingual aspect refers to the 'material' of which the poem is made. In turn, the aesthetic aspect refers to the type of normativity that determines its quality [2].

The ideas of philosophy of language demonstrate how a perspective in a scientific area might be developed to for incorporating life-meaning based methods into contemporary research.

The philosophical support points that language meaning is rooted in participating in human ways of life, and only human beings can function as subjects in the lingual and post-lingual aspects. What this means is that the human user needs to be involved intimately in the processing of text. 'I'll build a house right at eleven o'clock' is only understandable once one knows the life-meaning of communities that employ this idiom.

Despite some limitation, then it is likely that better rule sets can be compiled at least to detect and model the life-meaning issues with aim to improve the accuracy, namely the number of words correctly disambiguated, ipso facto preventing from word sense ambiguity, which is one of the main causes which affect the retrieval performance in the field of information retrieval and natural language processing, on account of the polysemies (when one word may have different meanings under different contexts) and synonymies (when different words may have the same meaning).

4 Nonverbal Aspects of Information Need

The concept of information need concerns the gap in a person's knowledge, urgently followed by craving for satiation of the adequate reference. The information request is expressed as a query asked for relevant sources.

An inquirer obtains an answer from an information system on a path of some levels of exploration. This starts with forming the actual, but still unexpressed need for information to evaluate then a rational statement and unambiguous description of the doubts and finally ask a question as presented to the information system.

To understand properly and respond adequately to users, the semantic extraction has become the major challenge in the development of the semantic web. Unavoidably, the complex process of extracting semantic information from natural language documents generates simultaneously particular difficulties. The semantic extraction from documents such as poetry is even more difficult due to its composite metric patterns, miscellaneous styles, ambiguity and elusiveness as well as symbolic meanings.

Arising from the philosophy of language, in consequence of the innate facility for language possessed by the human intellect, the natural language documents often appeal for nonverbal communication, as a means to reinforce and complement the message. Via its functions of complementing (adding extra information to the verbal message), contradicting (when our nonverbal messages contradict our verbal messages), repeating (used in order to emphasize or clarify the verbal message), regulating (serves to coordinate the verbal dialogue between people), substituting (occurs when a nonverbal message is transmitted in place of a verbal message) [11], the nonverbal communication is usually understood as the process of communication through sending and receiving wordless messages, however the written texts have also their nonverbal elements, such as spatial arrangement of words or emotional expression of feelings.

There are also areas where the verbal and nonverbal means of communication overlap, like in poetry, which carries rhyming patterns, rhythmic regularity in versification and metaphors with capability to express the inexpressible and assimilate dissimilarities. Thus, for poetry, with its metrical parallelism of lines and the phonic equivalence of rhyming words, metaphor, is the line of least resistance and, consequently, the study of poetical tropes is directed chiefly toward metaphor [10].

Traditionally, metaphor was a term in rhetoric, a term which referred to a purely linguistic figure, defined in the Aristotelian manner as the transfer of a name from one thing to another [4]. But with the advent of Structuralism, metaphor has acquired a greatly enlarged, granted that every realm of human culture can be construed as a type of language, what brings to light properties and relationships, which had hitherto gone unnoticed.

How far the nonverbal aspects of language can grasp and how deeply they can touch the ambiguity of the context, a simple task of text retrieval may visualize how it is used within linguistic context of a local culture in the Tatry region in Poland, a region of great natural beauty and folkloric originality, but completely novel in exploratory experience, including a peculiar local dialect, not readily comprehensible even to most of Poles. A locution 'I will build a house right at eleven o'clock' inclines towards a meaning of the time, when it will be raised, whereas the idiom practically concerns the building a house facing South. A user who wished to find texts that refers to the building a house facing South would completely miss those, that included 'To build a house right at eleven o'clock', unless the prevalence of retrieval methods respecting the nonverbal aspects of information need.

5 Structural Patterns of Poetic Works

The principle problem of the poetry retrieval is to find out the adequate methodology to define precise, relevant and accurate enough, web search model, able to handle with such obstacles like word sense ambiguity on account of the polysemies (when one word may have different meanings under different contexts) and synonymies (when different word may have the same meaning).

According to Chomsky's, the ability to learn and use language is an innate property, that generates the words in a hierarchical format of branched tree, called the structure of the syntax [5]. Chomsky's revolution shows that the internal capacity of learning, processing and building syntax is universal to the species, since we are born with sensitivity to this type of structural system, functioning on the principle of tree [15].

The nonverbal elements in real life can be, for instance, voice intonation, feelings expression by laughing, crying, sighing, as well as the elements of the physical

environment, like lighting, sounds, smells, whereas the nonverbal elements in poetry are expressed through stylistic means, such as onomatopoeic words, accent, rhythm, rhymes, expression of feelings, mood, metaphor, the spatial organization of words and thematic motifs as well.

As an example of versification pattern, appears alliteration, what is a repetition of the initial consonantal sounds of syllables and words, repeating the same letters and syllables at the beginning of words in the verse, the next verse or sentence. We can say that alliteration is like rhyming, but with alliteration the rhyming comes at the front of the words, instead of at the end.

As an example of rhythmic pattern, appears metric foot, what is a unit of measurement in poetry, while meter refers to the repeating pattern of stressed and unstressed syllables in the lines of a poem, just like in a screened example, where each unstressed syllable occurs in italics and each stressed syllable in bold.

The structural patterns of poetry are determined by, inter alia:

- literary genres: lyrics (ballad, epigram), epics (ode, elegy, pastoral, song, sonnet), drama (tragedy, comedy, monologue);
- literary eras: Middle Ages, romanticism, contemporary literature;
- mood (joy, sadness, happiness);
- thematic motifs (love, farewell, nature).

Application of methods for poetry retrieval, based on structural patterns of poetic works and adaptable linguistic models, will allow the efficient obtaining of relevant answers to questions, concerning the poems. Effective developing of the methods of poetry retrieval in web search can flow owing to adaptation of selected linguistic models on the structural patterns of poetic works.

6 Linguistic Model of N-gram Text Categorization

The Linguistic Model of N-gram Text Categorization is a probabilistic model that allows predicting the next elements in a sentence.

N-gram is a subset of sentences, composed of n-elements of the sentence, which may be letters, characters, syllables, words or pairs of elements mostly of sizes from 1 to 5. The most elementary uni-gram model rejects the conditional context and defines each term separately, hence usually used to find information in case of structural complexity documents, whereas the second size bi-gram model determines the occurrence of the words preceding [12].

The probability P(Wn|Wn-1) of Wn (where W is a word and n is a position in a sentence of Wn word occurrence), conditioned by the occurrence of a Wn-1 preceding word, is equal to the ratio of the probability of co-occurrence of two words P(Wn-1, Wn) to the probability of P(Wn-1) preceding words (equation 1).

$$P(W_n \mid W_{n-1}) = \frac{P(W_{n-1}, W_n)}{P(W_{n-1})} \tag{1}$$

The functions of distance between the document and category profiles are determined, according to the model, based on quantitative Zipf law, where word frequency is inversely proportional to the ranking, what means that the higher the rank, the lower frequency.

In data flow of the N-gram model of text categorization the following steps take place (Figure 1):

- text is divided into separate tags, consisting only of letters, an apostrophe and a space before and after the characters. Numbers and punctuation marks are discarded;
- each tag is scanned, generating all n-grams in size from 1 to 5;
- using the mechanism of collision each n-gram is given its own counter;
- all the n-grams are counted;
- N-grams are sorted in reverse order to the number of occurrences;
- the resulting file is a profile of n-gram frequencies for a document;
- the profile distances are measured;
- for each n-gram in a document profile, we find the equivalent in the profile category and count how far away. (For example, the n-gram "ING" is a ranking of 2 items in the document, but at 5 in the category, so it has a ranking of 3, after measuring the distance. If n-gram as "ED" does not appear in the profile category, with the value of the maximum distance.);
- the sum of all distance values for n-grams is a measure of the distance from the category of the document;
- to specify the minimum distance, the measure of distance for all profiles of categories to profiles of documents is taken and the smallest of them is a result.

The benefits of such approach is the ability to work both with short and long documents, also the minimum occupancy requirements of memory and computing, as well as a perfect fit to the texts of noisy sources.

The ability of predicting the next elements in a sentence through the process of measuring a distance between n-grams of the documents can also be effectively adapted to a language model for poetry retrieval, where the sum of all distance values for n-grams is a measure of the distance from the assumed words of the searched poems and the words of the retrieved poems, whereas the smallest of distances give an optimal result of poetry retrieval.

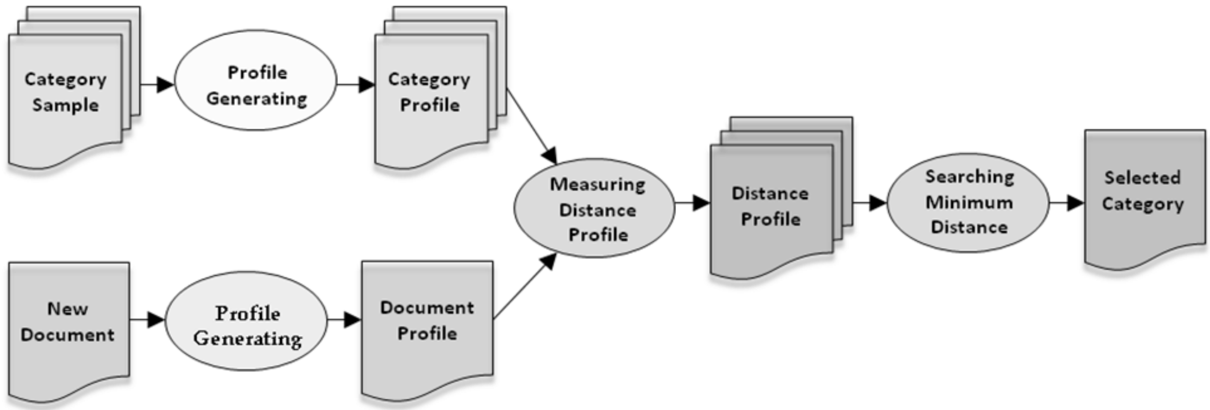

Figure 1: Data flow in the N-gram model of text categorization.

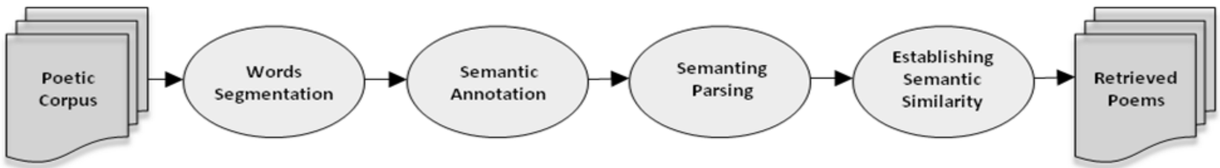

Figure 2: The flow of data in the semantic model of structural similarities.

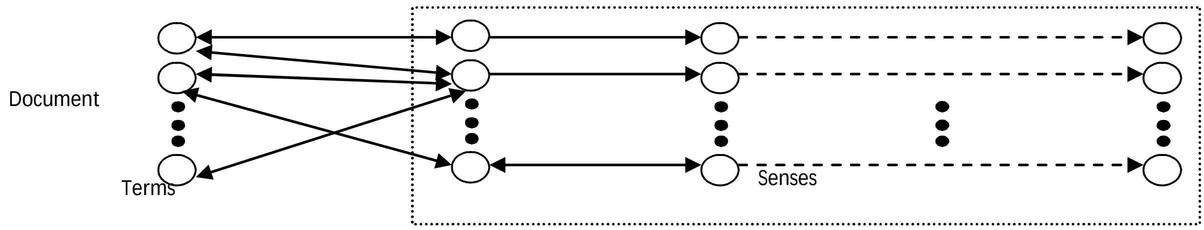

Figure 3: Language Sense Model for Information Retrieval.

7 Linguistic Semantic Model of Structural Similarities

Searching of poetry from the sources of poetic corpuses is done by measuring the scale of the semantic similarities on structural paths of queries and lines of poems on the basis of semantic annotation and parsing (Figure 2).

The Linguistic Semantic Model of Structural Similarities is based on the tree structure, adopted to search for poetry by using the algorithm of structural similarities [18], where:

- the T1 tree consists of n1 nodes;
- C(n1,n2) is the number of subtrees with roots at the nodes n1, n2 (When a semantic category for n1 is different than for n2, then C(n1,n2)=0, where n2, n1 are the final nodes and have the same category, meanwhile C(n1,n2)=sim(n1, n2).);
- K(T1,T2) is a measure of similarity between trees T1 and T2, which is the sum of similarity measure C(n1,n2) in each pair of nodes of trees T1 and T2 (equation 2);

$$K(T_1,T_2)=\sum_{\substack{n1\in N_1\\ n_2\in N_2}} C(n_1,n_2)$$

$$C(n_1,n_2)=sim(n_1,n_2) \quad (2)$$

$$sim(n_1,n_2)=\frac{1}{4^N}$$

- sim(n1,n2) – is the result of semantic similarity and capital N is a number of branches between n1 and n2 (In case, sim(n1,n2) is equal 1/16, then N equals 2, what means, that there are 2 branches between n1, n2; if sim(n1,n2) is equal = 1/256 then N equals 4, in other words, there are 4 branches between n1 and n2. Eventually, if sim(n1,n2) is equal = 1/4096 then N equals 6, and that means, there are 6 branches between n1,n2 and semantic similarity tends to infinity.

The flow of data in the semantic model of structural similarities shows that poetry retrieval in the sources of poetic corpuses is done by measuring the scale of the semantic similarities on structural paths of queries and lines of poems on the basis of semantic annotation and parsing.

The first step in the process is segmentation of words, when poetic lines of the body are segmented according to the dictionary of synonyms.

The next stage is semantic annotation, when expert annotates every word with the correct semantic code, in 4-leveled taxonomic hierarchy, in which the first level is the most abstract, and the fourth most detailed. Each term in a dictionary of synonyms has a semantic code that represents a hierarchical classification. For instance, for the word 'general' there is a semantic code AE1004, which is represented by four code elements AE 10 04, where A is the first-level category code for a person, AE means the profession of a persons, AE10 goes deeper for 'military rank' and finally AE1004 is a “specific name of military rank”.

Subsequently, each line of poems is being parsed on the basis of the metric pattern during the semantic parsing.

Eventually, the semantic similarity is being established through measuring the similarity of the semantic structure of each line of poems to the semantic structure of queries. Consequently, finding poetry is done by using the algorithm of structural similarities. The lower distance indicates the higher similarities.

8 Language Sense Models for Information Retrieval

Due to the vision of semantic web of universal flexibility to fulfil the requests of users, the extraction of the semantic information from natural language documents has become one of the major challenges to undertake. A great deal of work has been done on drawing word senses into retrieval to deal with the word sense ambiguity problem, but with few positive results. And yet the first significant accomplishments of such scientific achievement give the language sense model (LSM).

A document model generating queries can be used either to recognize or to generate strings. The full set of strings that can be generated is called the language of the automaton, consisting of nodes, capable to exchange information over a communications channel. If each node has a probability distribution over generating different terms, then we have a language model [12], which is a function that puts a probability measure over strings drawn from some vocabulary.

In the LSM the model generates the probability of a given query from both document’s term and sense representation (Figure 3).

he LSM combines the terms and senses of a document seamlessly through an expectation-maximization (EM) algorithm for data augmentation and optimizing a likelihood function, what lets overcome such problems of research work related to language model as data sparseness (the existing smoothing methods can be applied easily on both terms and senses to solve the data sparseness problem) and term dependency (the query independent assumption can be relaxed to a certain extent as the terms and senses in the LSM depend strongly on each other) [1] (equation 4).

$$P(q \mid d) = \prod_{i=1}^{m} ((1-\lambda)P(q_{t_i} \mid d_t) + \lambda P(q_{s_i} \mid d_s)) \quad (4)$$

Retrieval on lexical collections shows that the LSM outperforms the traditional language model for both medium and long queries, however, not significantly on short queries, as there are less nouns and verbs to be disambiguated for short queries as well as it’s much harder to disambiguate the short queries because of the sparse context.

The forthcoming attempts anticipate the hierarchical smoothing using more relations of lingual database groups words into sets of synonyms with semantic relations between them.

The purpose is to produce a combination of dictionary and thesaurus that is more intuitively usable, and to support automatic text analysis and artificial intelligence applications throughout the evaluation of the LSM on more corpora.

9 Conclusions

Usage of poetry retrieval methods, based on structural patterns of poetic works and adaptable linguistic models, including the distance measure of words and phrases similarities about non-verbal meaning as well as building new linguistic models and using structural patterns of poems, will allow the efficient acquisition of response to questions, which concern poetry retrieval.

Exploring the websites databases of poetry works, based on linguistic models, shows that the present systems of poetry retrieval reveal weaknesses within identification of the contextual ambiguity and under accuracy of information retrieval.

Validity of information retrieval in context of non-verbal requires of users emphasizes this year's report by the

International Telecommunications Union, which shows that yet 10 years in the whole world was less than 400 million Internet users, while today there are exactly 2.08 billion, which means that about 1/3 of humanity benefits from Internet resources, including the 950 million users of mobile broadband, and 555 million from fixed broadband with a steady upward trend of potential target groups and relative areas of application, carrying the indisputable benefits of thorough exploration to personal growth and flowering of modern education and information society.

It can be concluded that poetry retrieval might be both possible and beneficial, giving opportunity of effective efforts towards thoroughgoing web collection research.

10 References

[1] Bao S., Zhang L., Chen E., Long M., Li R., Yu Y. (2006). LSM: Language Sense Model for Information Retrieval, Lecture Notes in Computer Science, pp. 97-108.

[2] Basden A., Klein H. K. (2008). New research directions for data and knowledge engineering: A philosophy of language approach Introduction to Information Retrieval, ELSEVIER/Data & Knowledge Engineering, Vol. 67, No. 2, pp. 260–285.

[3] Bojar B. (2002); Słownik encyklopedyczny terminologii z zakresu języków i systemów informacyjno-wyszukiwawczych.

[4] Bredin H. (1990), Metaphor, Literalism, and the Non-Verbal Arts, Springer Netherlands, Vol. 20, No 3, pp. 243-341.
[5] Chomsky N. (2000). New Horizons in the Study of Language and Mind. Cambridge University Press, pp. 1-195.
[6] Dooyeweerd H. (1984). A new critique of theoretical thought. Paideia Press, Vol. 1-4, pp. 1-2205.
[7] Donald R. B. (2000). Nonverbal Poetry: Family Life-Space Diagrams; Journal of Poetry Therapy, Vol. 14, No. 3, pp. 159-167.
[8] V. Fomichov and O. Fomichova (2006). Cognitonics as a New Science and Its Significance for Informatics and Information Society. Special Issue on Developing Creativity and Broad Mental Outlook in the Information Society (Guest Editor Vladimir Fomichov), Informatica. (Slovenia), 30 (4), pp. 387-398.
[9] O. Fomichova and V. Fomichov (2009). Cognitonics as an Answer to the Challenge of Time. Proceedings of the International Multiconference Information Society- IS 2009, Slovenia, Ljubljana, 12 – 16 October 2009. The Conference Kognitonika/Cognitonics. Jozef Stefan Institute, 2009, pp. 431-434; available online at http://is.ijs.si/is/is2009/zborniki.asp?lang=eng.
[10] Jakobson R., Halle M. (1956). Fundamentals of Language. The Hague & Paris: Mouton, pp. 55-82.
[11] Malandro, L. A, Barker L. L., Barker D. A (1989). Nonverbal Communication. Reading MA: Addison-Wesley.
[12] Manning C. D., Raghavan P., Schütze H. (2008). An Introduction to Information Retrieval; Cambridge University Press, pp. 237 –252.
[13] Marchionini G., White R. W. (2009), Information-Seeking Support Systems, IEEE/Computer, Vol. 42, No. 3 (pp. 30-32).
[14] Maslow A.H. (1943). A Theory of Human Motivation; Psychological Review, Vol. 50, No. 4.
[15] McWhorter J. (2008). Understanding Linguistic: The Scence of Language; The Teaching Company.
[16] Mehrabian, A. (2002). Silent messages. Nonverbal communication; Reedition Publishing, Springer-Verlag, New York.
[17] Reitz J. (2004). Dictionary for Library and Information Science; Libraries Unlimited, pp. 357.
[18] Shu-Lei Chen W. (2004). Semantic Structure Extraction and Retrieval of Chinese Poetry, National Tsing Hua University, pp. 1-49.

The Experiences of Landscape Social Perception as a Remedy for Plunging into Virtual Reality

Rita Micarelli
International Institute for Advanced Studies in Systems Research and Cybernetics (IIAS), Italia
E-mail: rita.micarelli@libero.it

Giorgio Pizziolo
University of Florence and IIAS, Italia
E-mail: pizziolo@unifi.it

Keywords: social learning, cognitonics, landscape, virtual/natural experiences, prevention of plunging into virtual reality

Received: January 6, 2012

A progressive exasperation affects the relation men/society/environment and puts at risk its very evolutionary potentialities. At the same time the increasing influence on natural world by the virtual globalized powers accents and exasperates their separation and their reciprocal communications. Hence a new way of learning is needed beingable to re-constitute experiential modalities and encourage the reciprocal integration between these two conditions. A new idea of the life environment – virtual and natural - can be suggested for a concrete and socially practicable learning based on a multiplicity of landscaping research-action, being able to bring to light unexpected evolutionary learning contexts, where Landscape becomes the interconnection's hinge between territory/society Entities and virtual /experimental Minds. New procedures based on the relationships between social groups and Life environments can encourage effective research-actions on different contexts to harmonize social learning and landscape practices in a continuous virtual and real entwining and to prevent young people from plunging into virtual reality.

Povzetek: Prispevek se ukvarja z relacijo človek/združba/okolje in predlaga nove postopke v izogib ponoru v virtualno realnost.

1 Introduction

The contemporary world, with its life environments, is characterized by two (apparently opposite) phenomena: the first is related to the immaterial informational network, wrapped - like a *labyrinth-grid* - around the planet, the second affects the real life environments where nature and human activities interact, in a reciprocal, continuous transformation.

Both these phenomena suffer from their intrinsic and reciprocal contradictions towards the social learning processes: the *virtual* one because it tends to exasperate the learning processes (fast, quantitative, short, placeless), the *human/natural* one because it tends to isolate the natural places of the social learning processes (jeopardized places, isolated communities, subsistent level of economies and cultures).

The learning processes which occur in these two worlds are more and more separated and divergent, both at risk of implosion. At the same time, a lot of conflictual tensions are increasing between them. In this situation the *distortion of the perception* of the life environment and the creation of *new cognitive-cultural pre-conditions* referred to landscape experiences can be developed and discussed within the new scientific discipline called Cognitonics [3, 4].

This paper aims at implementing these complex issues, on one hand, by means of going into this matter more thoroughly to discover the learning processes intrinsic to the living world and, on the other hand, by means of discussing these processes in relation to the digital dynamics of the virtual word.

2 The Idea of Using Social Learning Processes for Preventing the Plunging into Virtual Reality

The contradictions described above have been progressively accented in the recent years. Actually many people, especially the young contemporary generations, risk to lose the perceptive experiences towards their life environments together with their spontaneous social relationships. Hence, they tend to emphasize only the virtual - instantaneous superficial - interactions.

However, numerous observations accumulated in various countries have shown that plunging into the virtual life prevents people (especially young people) from solving real life problems, from building social relations in the surrounding world and leads to breaking family ties and other social ties. The person is unsuccessful in solving the problems of real life, when some external circumstances force he/she to do it.

Therefore, the urgent, crucial question is now how to find the ways of using the social/life environment contexts for helping people (especially young people) to appreciate the real world around them and to not plunge into the virtual life. In other words, the urgent question is how to address the *social/life environment contexts* toward appropriate learning experiences ?

The traditional educational procedures can't cope with this unstoppable, frenzy problem. Nevertheless, we could think that some attracting landscape experience would be welcome even among these *absent* and *placeless groups*.

Same study-examples (already texted) can be helpful to understand the meaning and the *methodological effectiveness* towards the adult and child groups, the living contexts and institutions progressively involved. Different social groups and living contexts in North/Central Italy participated in such activities during the last twenty years. Their processes can be valued under multiple points of view (procedures, cyclical balances, evolutionary trends) referred to the re-conquered self consciousness by the groups through the re-discovery of their life environments.

These researches selected among a lot of case studies can be divided in two groups:

Group 1: Landscape art, friendly participation, social awareness, didactic experiences.

Group 2: Explorations among users and inhabitants of contemporary *virtual landscapes* and *real life environments,* encouraging their participative meetings.

It should be underlined that the experience obtained with these two groups goes beyond any traditional political-planning or project methods.

Specific narrative and figurative art languages, addressed to young, child and adult people, were used during these processing-experiences in order to improve reciprocal confidence and relationships among all participants. These procedures encouraged both aesthetic appreciation of life environment and further friendly relationships among all participants. These three experiences were tested between the 1990's and 2004 involving social and child groups [5, 6].

Figure 1: A social Group at work, preparing a landscape Contract.

3 The Experience with the First Group "Landscape- Art, Friendly Participation, Social Awareness, Didactic Experiences"

This first group includes three experiences developed on various territorial contexts with several different social groups (adults, children) invited and spontaneously disposed to these activities, i*nvolved in several creative processes* concerning *social perception of landscape*. A variety of friendly meetings, spontaneous conversations, and promenades guided by people throughout their life spaces have been experienced. In general the knowledge and the perception have ranged from the widest landscapes to the smallest public spaces selected *in loco*.

Such activities have been developed as *process-experiences*, progressively and chorally implemented by experts and people (meeting, discussing and learning together) in a pleasant atmosphere, in a very 'friendly learning'. All of these experiences allowed experts and people to achieve a new appreciation of their own life environment, and some time they opened up also new opportunities of social creative participation in its transformations.

3.1 Case study 1 (Suvereto)

The birth of our methodological approach "Social Perception of the Landscape" is associated with the experience obtained in Suvereto Village, south Tuscany, Province of Livorno (1995-98)[1].

The promoters: University of Florence, National Research Council

The research aims: Relationships between the population and the territory, working through the Landscape in a participative planning perspective, from the landscape mosaics to Landscape holistic social perception and evaluation

The participants: scientific consultants, social, scholar and institutional groups, all directly involved to obtain answers to our questions.

The procedure we have followed:

[1] The coordinator of this research is Giorgio Pizziolo. Rita Micarelli and Paola Paradisi took part in the project experience as well, the first one for the social ecology and the latter for the psychosocial investigation.

- we involved every subject in the knowledge process; accepted as a very intriguing play by the participants.
- we focussed the discussions, on “landscape ambits”, with the explicit intent of recognise those ambits on the base of the differences among them, in terms of living experiences.
- then we asked a personal evaluation to the subjects that participated in the experiment, through different modalities (drawing and description for the elementary school children, proper and more specific questions for the junior high school students, and articulated questions for adults who had been provided with direct answers in “classroom meetings”).
- we examined the collected material, working out different hypothesis of interpretation.

Finally we have promoted a new meeting to discuss with the people who had been interviewed the results of our research and the validity or the lack of validity of the suggested hypothesis, and we accomplished a balance of the whole operation and its possible developments.

All the subjects knew perfectly well that the environment in which they spent their every day’s life, with different times and modalities, according to their age groups, was *their own Oikos*. Such an *Oikos* was intended as the whole, the everything – straight perceivable and checkable-, which could become such though the intensity of the relations established between the “inhabitants and the inhabited place”.

The quality of the perception and the valuations that everybody was showing out, and expressed by different languages:

- The **child** expressed their critical opinions or positive appreciations not by a formal answers a questionnaire but throughout integrated activities, surveys on the places, informal colloquies, sketches, coloured drawings, discussed and valued by the scholar groups together with the experts and their teachers, as a very didactic activity.
- The **eldest group** pointed out some marks in the landscape, invisible to others, of past relations, of frequentations, and activities they perceived as their own, yet socially shared. (the habitual frequentations on the territory, the “reconnaissance walking” in the woods the territory maintenance, the entertainment activities)
- The **young people**, even studying or working elsewhere, remained deeply rooted and strongly related to their original life environment, recognised as a sort of privilege, to respect and save jealously, as a firm point in their lives.

All of the participants (included the Major) shared the hypothesis of a territory considered as an Oikos, showing a significant social and cultural cohesion of this people.

The inhabitants themselves have considered this process to be so involving that they wanted to continue the activity, and they proposed to go towards a new hypothesis, moving from the “Social Perception of the Landscape” to the “Social Managing of the Landscape”, and therefore find a way to guarantee its use and active preservation through a participated maintenance.

3.2 Case study 2 (Vezzano)

Our second case study concerns the Vezzano’s participative landscape experiences: villages, agriculture, waters and… *enjoyed panoramas* ! (1998-2002, La Spezia Province, Liguria Region, North Italy.

Vezzano is a small village of the Liguria’s Rising Sun Coast, at the confluence of Vara and Magra Rivers, and overlooking the more distant Gulf of La Spezia, the Apennine roads, and the Apuane Alp passes, in ancient times exchange and sentinel crossroad for goods, abundantly rich in fruits, wines, oil, gardens, inhabited by noble families living in their palaces and by poor, yet not miserable, farmers. The village is a blaze of colors, blue, green, and grays among waters, stones and the sky. But nowadays Vezzano is risking to be choked with highway junctions, the container storage areas of La Spezia harbor, the dried areas of the polluted rivers, just like the outskirts of La Spezia, city that in the last thirty years has been piling up poisons, traffic jams, and pollution, without getting going toward the wonderful and great destinies of the Gulf.

Against these macro- technological interventions and for proposing micro environmental new interventions, the itizens REACT, taking care of their beautiful life environment, starting a new landscaping action:

- The landscape is carefully maintained and properly used by scholar groups, citizens and local farmers, to promote friendly touristic reception and micro-economic activities.
- Ancient fountains, profoundly and affectionately linked to local memories, had been abandoned. A group of citizens re-discovered and restored all of them with a spontaneous project promoted by inhabitants.
- This attracting landscaping action began some years ago and is still successfully practised also in other places of local landscape

3.3 Case study 3 (Pescia)

Our third case study can be named as “Art, Sociality, and Science walk in a Mutual and Friendly Learning process” (Pescia, North Tuscany, a participative research-action).

This activity became an opportunity to experiment a new method of learning, through a cognitive and creative process focussed on the relationships among inhabitants and their life environment

The authentic friendly learning process has been progressively built up by progressive friendly meetings between experts and motivated groups of people to develop *new relations with their life environment.*

In this process each group expressed its own valuations, desires and options (social problems, urban renovations, a developing program, a political disagreement, a special need, etc.) and discussed about (new schedule, suggestions, proposals, further reflections, and theme choices for the creative participation to the transformation of one's own context) for a further developments for the process in cyclical verifications and creative shared perspectives.

In particular the relationships between School and Territory encouraged new participative landscape interactions.

These interaction have been experimented and developed through participative procedures, appropriate to the anthropological, social, and territorial condition of Pescia, a small town so much differentiated in its varied territorial and social "zones". Such a procedure allowed us, to get to a program of **re-qualification of the life environment of relational kind**, by developing the great proposal qualities contained in the "**relational systems**".

4 The Experience with the Second Group

As it was explained above, the Group 2 is formed by explorations among users and inhabitants of contemporary virtual landscapes and real life environments, encouraging their participative meetings.

This second group of experiences began only *recently* and at the present is in progress. We are trying to involve some group of young *web professional and users* of the informatics networks, often (even if latently) conscious of the intrinsic limits of such so potent instrument towards their personal concrete experiential potentialities.

The hypothesis is to create new concrete opportunities of reciprocal exchanges between *widespread virtual communications* and *cultural life environment phenomena*.

According to this initial hypothesis, we propose '*new contacts' modalities* which affect at the same time individuals and groups, together with their landscape references (both in concrete life environment terms and in *virtual landscape* terms). All contacts must be focussed on *thematic meetings* and concrete *lasting process perspectives*. These meetings can produce different interesting results both towards *young and child people*, wide-ranging from didactic experiential level to real opportunities in terms of social knowledge and sustainable economies based on the initially characterized landscapes.

A perspective for the second group is given by the following procedural and methodological implementation hypotheses:

2a *The integrated economies of the Commons*

2b *The micro innovative or restored activities towards mountain economies*

2c *The creation of new social managements of urban life environments as cultural concrete opportunities to go beyond the social exasperate protests and legal claims*

2d *The creation of a child unexpected Landscaping vision* (scholar experience on the water front of la Spezia Gulf, Portovenere, the U.F.O. land on the Park)

Figure 2: A Brownian, aimless promenade.

5 A proposal for an Epistemological Effective Discussion

Before dealing with the *social life environment* processes, we have to consider more in general the learning processes, as based on the relationships *individual/social/ life environment*, and referred to the fundamental evolutionary structure defined by Gregory Bateson as the *Ecology of the Mind and the Nature* [1, 2].

5.1 An approach of G. Batson to ecology

This renewed idea of the Ecology constitutes an extraordinary epistemological suggestion that may be has not been adequately considered in its widest significance, just towards the individual and social behaviors, at the present more and more influenced by the "digital planetary powers".

G. Bateson held in his books [1, 2] the idea that the mental processes could be influenced by a double behavior, digital and analogical at the same time. In particular he focused the working of the neuronal communication as a *double interaction* which ranges between quantitative and qualitative ways of communication. Actually, according to the suggestions arose from the *Criteria of the Mental Processes* defined by G. Bateson[2], the *response* to the *digital* stimulus (*yes or not)* and the *analogical response* to the *ostensive communications* are never completely separable and they remain both the intrinsic factors for all of the *living learning processes*.

[2] The IV chapter of *Mind and Nature* is dedicated to the six "Criteria of the Mental Processes"

This interactive co-existence remains still a crucial question for our contemporary life, where the digital ways are more and more pervading the social behaviors, especially towards the mental and material life environments, both progressively dominated by the *digital planetary processes.*

In this sense, it seems once more very urgent to us, to re-consider and assume the epistemological *Criteria and Principles* as proposed by G. Bateson whenever we have to do with individual and social learning processes, just while the dominant powers tend to control *even* the roots of all these processes in the world.

At any level of such processes the role of the experiential factors (analogically consequent to ostensive communication) has been progressively reduced. In this way these processes risk losing their balanced equilibrium and their evolutionary attitudes.

The *Steps to an Ecology of Mind* achieved by G. Bateson lead us to a more complex epistemological level (as a new state of the knowledge) which goes beyond the exasperated conditions above mentioned, running toward new relationships between Mind and Context, meant as two homologous entities within which the learning processes happen. In this operating way the Contexts become a *Locus* and –similarly- the Mind consolidates its quintessence of *ambit of processes*. And at last a resonance can be established between *Locus* and *Mind,* throughout qualitative and quantitative stimulus from which homologous processes are generated.

5.2 How to develop the *learning dynamics* in coherence to the Ecology of Mind in our contemporary conditions?

To answer this crucial question, we have to explore in depth the modalities of the social learning related to the contemporary life environments.

We assume, on one hand, the *life environments* as different fields where human societies practice analogical experiences (social and individual) and, on the other hand, the widest *web environment* as the place where unlimited digital explorations are possible. These conditions are often perceived as two incommunicable realities- often separated – without reaching an adequate level of profitable reciprocal interaction. Within these opposite worlds individual and social pathologies are progressively manifesting and become more and more exasperate.

On one side we meet different micro realities, rich in complexity but firmly rooted and bound in their living contexts, that can ramify within the interstitial spaces and the marginal phenomena. Meanwhile, on the other side, we are faced with an infinite grid, rich in differences but bound in its rigid digital consumeristic dynamics. In these conditions neither the first nor the second world are able to become a very contemporary *Locus* where new evolutionary learning processes could be developed according to the principles suggested by the Bateson's *Ecology of Mind*.

The two sceneries seem still to be in a paradoxical situation that could be passed only through a new epistemological approach.

In this regard it would be possible to consider these two separated sceneries as two components of a *whole learning processes of the Mind*, as suggested by Bateson. These sceneries can be likened to the homologous phenomena that happen at any level of the living world, from the *micro biological systems* to the *widest spatial and social dimensions*. In this way it becomes possible to connect these two worlds throughout a new kind of *relational interaction* that, as it happens in nature, creates new tensions among the *digital* and *analogical* respective dynamics.

In this contemporary age in which the globalized *mono cultural* thought and the *one- model* modalities (economical, scientific, behavioral) are prevailing, how the micro living realities can resist?

They can do it only by using the well known technologies of the virtual communication as the recent Arabian, European and American protesting phenomena have been showing. Nevertheless such instruments can be profitable not only to help social protests but also for unexpected innovations propagable among social groups (young people in particular). In this sense these innovations could expand their interstitial ramifications and establish a profitable connection to selected websites. In this way the local interstitial structures could widely spread and, at the same time, a new virtual relational structure could come to light, as a new *context of the flowing communications*.

Thus two new kinds of *Locus* come to view: the *Locus of interstitial context experiences* and the wider *Locus of virtual/relational interactions.*

Each *Locus* is related to its community which take care of it by a different specific manner: the local community throughout a sustainable management, the virtual community by a continuous implementation of information's inter-exchanges.

5.3 The possibility of emergence of a new evolutionary complexity

In this sense, according to the mentioned extraordinary epistemological suggestions, we have *at first* experienced various modalities of this learning, developed in a relational sense between life environments and social contexts, and *later* we proposed new relational ways to connect the *virtual* and *real* life environments both referred to contemporary realities.

We have experienced many processes where the social, environmental expert and professional subjects worked in synergy to experience in concrete a reciprocal friendly learning. These early experiences allowed us to approach further participative dynamics within the virtual/real landscape contexts.

The following question can be put forward: how to experience the evolutionary meetings of virtual and real contexts? We mean here the contexts where such meetings could be promoted by a relational extended idea of Common Good (from the classic territorial Commons to the Informatics' Networks as suggested by E. Ostrom). In this way reciprocal wider exchanges, not generically approached but focussed on *thematic terms* (informational, didactics, scholar experiences, professional collaborations, experiential procedures, comparability, mutual specific supports, autonomous economies), could become the new structure able to involve persons and groups coming from different conditions (cultures, lifestyles, economies, life environments) brought in contact by such integrated (virtual and concrete) holistic modalities.

The European Landscape Convention remains a basic reference for all these new activities and for the promotion of a new way to deal with the life environment for the involved populations.

Hence, no more classic methods, no more models to apply, but rather different criteria and wider approaches, sensitive and participated observation of the phenomena, informational network's interexchange, could be helpful to constitute unusual **real and non hierarchic virtual groups**- experts, citizens and users who work together in a new, **extended life environment.**

In such environment experiments and researches could be constantly and mutually compared, in an open cyclic progression. In this perspective the virtual and the living contexts (with their users and inhabitants) could be widely referable to the relational principles suggested by G. Bateson in *Mind and Nature.* This *context* could become a new, unexpected **Landscape,** where the Nature, the Mind and the non hierarchic Informational Network equally interact.

We can ask how is it possible to *GRASP* such opportunities and what does it really mean the word GRASP? We can read this word ***GRASP*** as following**:** **Groups** of **Research Action** for **Solidarity** and **Participation.** Consequently we can describe more in detail, *who, what, where, when, why* these groups come to light and work**.**

Who can participate in the GRASP activities**:** all of the persons, professional, cultural and scientific institutions, associations, local communities, with their competences, experiences, desires, aimed to develop Researches and Actions dedicated to the construction of new ways to improve the social quality of life.

What are the activities of these groups**:** promotion of social action towards local and territorial economies (micro industry, agriculture, culture) to encourage multiple relationships, free circulation and integrate sympathetic interexchange.

Where such activities can be developed: in every place within which a new relation between living context and inhabitants comes to light (rural environments, urban and metropolitan peripheries, abandoned areas, obsolete infrastructures monumental landscapes and places).

When all of these activities manifest: whenever the crises or significant transformations emerge (at economical, social and cultural levels).

Why it is necessary to assume this relational /informational dimension: because the complex contemporary situation requires appropriate places to open out its positive potentialities, like new real/virtual *LOCI.*

The five points above listed can lead us to appreciate the difference between the *GRASP* and the myriad of network, spontaneous local groups, already existing and working.

The *GRASP* are not a mere multidisciplinary aggregation of competent subjects or groups. On the contrary they assume the feature of a new *complex whole organism*, able to *conceive* and *produce steady equilibrium and dynamic transformations* both towards themselves and the external environments. In this way they are *going beyond* the multiplicity of exasperate motions of the network already existent (which evoke the Brownian aimless motions) and the local self bounded experiences.

Then the GRASP feature themselves as connecting structures endowed with an intrinsic complex relationality.

In this sense the *GRASP* can play a key role of promoters of new experimental interexchange and solidarity activities, becoming a fundamental interface between the informational network and the level of the micro local working subjects.

In this way a new kind of dynamical and participative connection between the informational and real worlds can originate a new complex social learning, which, by a fluent motion, can freely circulate among the (macro) virtual and (micro) natural conditions of this contemporary age.[3]

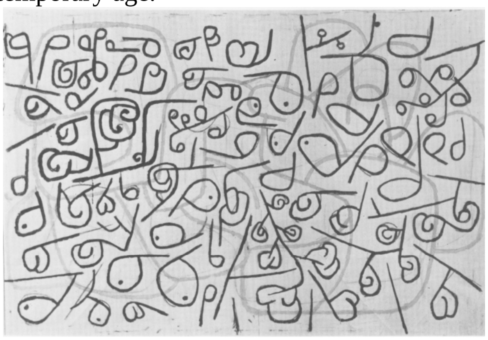

Figure 3: Paul Klee*,* Without title, 1940.

6 Conclusion

The contraposition between virtual and experiential life which affects the contemporary societies can be concretely approached going beyond their current separate condition, toward a wider challenge based on participative processes, equally referred to the contemporary *local contexts* (as landscapes/life environments) and to the *virtual* environments. In this sense the landscapes become the living contexts where the local communities can experience specific *analogical* learning processes (social perception, consciousness,

3 At the present the promotion of the GRASP is in progress in Italy, among social spontaneous aggregations already working both on informational and on real places.

creativity) while the virtual environments become the places where different experiential processes meet by an *equal digital* interchange.

A new *relational environment* can come to light as these two modalities of interchange move together and interact to originate more and more *complex learning processes*.

The researches we have practiced until now demonstrate that *a new* evolutionary coexistence between contemporary societies and life environments is possible and can be realized in the course of learning processes, *digital* and *analogical, virtual* and *material* at the same time. Due to these procedures, new relational life environments can come to light as the contemporary *Loci* for the *Ecology of Mind and Nature.*

7 Acknowledgement

Our special acknowledgements are dedicated to the friend, members and professors of The International Institute for Advanced Studies in Systems Research and Cybernetics (IIAS): G. Lasker, A. Aydin, K. Hivaki, K. Boullart,, H. Schinzel, J.R.L. Chandler. M. Malatesta, A. Grappone, S. Santoli, V. Fomichov and O. Fomichova. We have been discussing with them our hypotheses during the IIAS InterSymp conferences, which are annually held in Baden-Baden (Germany).

References

[1] G. Bateson (1979). *Mind and Nature*. Publisher G. Bateson.

[2] G. Bateson (1991). *A Sacred Unit, Further Steps to an Ecology of Mind.* Publisher G. Bateson.

[3] V. Fomichov and O. Fomichova (2006). Cognitonics as a New Science and Its Significance for Informatics and Information Society. *Special Issue on Developing Creativity and Broad Mental Outlook in the Information Society (Guest Editor Vladimir Fomichov), Informatica. (Slovenia)*, 30 (4), pp. 387-398.

[4] V. Fomichov and O. Fomichova (2011). Preface/Predgovor. Second International Conference on Cognitonics (Cognit 2011). In *Proceedings of the International Multiconference Information Society – IS 2011, Slovenia, Ljubljana, 10 – 14 October 2011, Vol. A*. Jozef Stefan Institute, 2011, pp. 349-350; available online at http://is.ijs.si/is/is2011/zborniki.asp?lang=eng.

[5] R. Micarelli and G. Pizziolo (2003a). *L'arte delle Relazioni* (*the Art of Relations).* Alinea, Firenze.

[6] R. Micarelli and G. Pizziolo (2003b). Dai margini del Caos, l'ecologia del progettare. Alinea, Firenze (From the margins of Chaos, the ecology of creating project).

Using M Tree Data Structure as Unsupervised Classification Method

Marian Cristian Mihăescu and Dumitru Dan Burdescu
University of Craiova, Faculty of Automation Computers and Electronics, Romania
E-mail: mihaescu@software.ucv.ro, burdescu@software.ucv.ro

Keywords: M Tree, k-means, unsupervised classification, cognitonics

Received: December 8, 2011

Increasing the effectiveness of educational processes is one of the greatest challenges for information society. The paper presents the usage of M Tree structure for classification of the learners based on their final marks obtained in their respective courses. The classical building algorithm of M-Trees with an original accustomed clustering procedure was implemented. The data that are managed within M Tree structure are represented by instances. The main goal of the structure is to provide information to students and course managers regarding the knowledge level reached by students. The proposed clustering procedure that is used for splitting full M Tree nodes is designed to properly classify learners. A baseline classification scheme based on k-means clustering and a custom M Tree clustering are presented. For comparison, there are considered classical characterization formulas.

Povzetek: Opisana je metoda za izboljšanje učenja na osnovi M-dreves.

1 Introduction

The ability to classify a student's performance is very important in internet-based educational environments. A very promising area to attain this objective is the use of special designed data structure. In fact, one of the most useful applications of modern algorithms in e-Learning is classification. E-students are students that follow courses within an e-Learning platform. There are different educational objectives for using classification, such as: to discover potential student groups with similar characteristics and reactions to a particular learning strategy, to improve a student's capacity of learning, to group students who are failure-driven and help them improve their skills, to identify learners with low motivation and find remedial actions to lower drop-out rates etc. In the followings we have applied a classification method using unique algorithms which have a common base (tree classification).

One of the greatest challenges in e-Learning area is to continuously improve existing systems. In order to overcome the challenge there are needed sound procedures whose task is to prove the challenger procedure creates a better system than existing one. The key issue regarding effectiveness of educational process is classification. The main goal of this paper is to obtain a better or at least acceptable classification scheme with less computation power.

Some possible outcomes of such analysis process are: predicting students' grades (to classify in three classes/clusters of low priority – week learners, medium priority – easy learners, high priority – competitive learners) from test scores.

Clustering algorithms are part of the unsupervised classification techniques. They try to group a set of items into subsets or clusters. The cluster algorithms' goal is to create clusters that are coherent internally, but clearly different from each other. In other words, items within a cluster should be as similar as possible; and items in one cluster should be as dissimilar as possible from items in other clusters. In this paper, learners represent items.

The standard k-means algorithm [1] is used as baseline unsupervised classifier. K-means is the most important flat clustering algorithm. Its objective is to minimize the average squared Euclidean distance of items from their cluster centres where a cluster centre is defined as the mean or centroid of the items in a cluster.

M-tree [2,3] is a dynamic access method suitable to index generic "metric spaces", where the function used to compute the distance between any two objects satisfies the positivity, symmetry, and triangle inequality postulates. The M-tree design fulfils typical requirements of multimedia applications, where objects are indexed using complex features, and similarity queries can require application of time-consuming distance functions. In this paper we describe the basic search and management algorithms of M-tree, introduce several heuristic split policies, and experimentally evaluate them, considering both I/O and CPU costs. The obtained results also show that M-tree performs better than R*-tree on high-dimensional vector spaces.

2 Related Work

It is now recognized that e-learning further requires the means to summarize and classify learner trends and patterns. One serious candidate solution is DM (data mining), already quite successful in e-commerce and bio-informatics, where results are achieved through the use of associates, classifiers, clusters [8], pattern analysers, and statistical tools.

Educational Data Mining [7] is an emerging discipline concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in.

Since the mid-1990's, e-learning has epitomized a broad range of learning categories while reinforcing four major pedagogical perspectives often neglected during e-learning system development. First, insight from cognitive learning processes can shed light on how the brain functions. Second, emotional aspects of learning can be traced, such as interest, motivation, interaction, fulfilment, and enjoyment. The third perspective incorporates skills and behaviours, such as role-playing, that are particularly useful in real settings.

Lastly, a social perspective involving the interaction with other people permits a focus on collaborative discovery, namely, the interplay of peer pressure and support. The complexity of the approach is high, and the specialists have been more preoccupied about the development of the information systems from the perspective of the technological informatics infrastructure. The studies devoted to the technology infrastructures embedded in the information systems are insufficiently presented in literature [5].

One of the constructive steps in this direction was done by V. Fomichov and O. Fomichova in [6]. The authors introduced the notation of conceptual-visual dynamic schemes (CVD-schemes). The CVD-schemes are the marked oriented graphs introduced in cognitonics domain for inventing effective teaching analogies. Such graphs establish a correspondence between the components of a piece of theoretical material to be studied and the components of a well-known or just created by the teacher but bright fragment of the inner world's picture of the learner.

Novel database applications, such as multimedia, data mining, e-commerce, and many others, make intensive use of similarity queries [2] in order to retrieve the objects that better fit a user request. Since the effectiveness of such queries improves when the user is allowed to personalize the similarity criterion according to which database objects are evaluated and ranked, the development of access methods being able to efficiently support user-defined similarity queries becomes a basic requirement. In this article we introduce the method called the M-tree that can process user-defined queries in generic metric spaces, that is, where the only information about indexed objects is their relative distances. The M-tree is a metric access method that can deal with several distinct distances at a time: (1) a query (user-defined) distance, (2) an index distance (used to build the tree), and (3) a comparison (approximate) distance (used to quickly discard from the search uninteresting parts of the tree). We develop an analytical cost model that accurately characterizes the performance of the M-tree and validate such model through extensive experimentation on real metric data sets. In particular, our analysis is able to predict the best evaluation strategy (i.e., which distances to use) under a variety of configurations, by properly taking into account relevant factors such as the distribution of distances, the cost of computing distances, and the actual index structure.

The access method called M-tree is proposed to organize and search large data sets from a generic "metric space", i.e. where object proximity is only defined by a distance function satisfying the positivity, symmetry, and triangle inequality postulates. The M-tree design has been motivated by retrieval requirements from typical multimedia database applications, where objects, such as text, image, and video, are indexed using complex feature representations, and search for objects similar to a query object can involve application of time-consuming distance functions. We detail algorithms for insertion of objects and split management which keep the M-tree always balanced - several heuristic split alternatives are considered and experimentally evaluated. Algorithms for similarity (range and k-nearest neighbours) queries are also described. The results from extensive experimentation with a prototype system are reported, considering as the performance criteria the number of page I/O's and the number of distance computations. The results demonstrate that the M-tree indeed extends the domain of applicability beyond the traditional vector spaces, performs reasonably well in high-dimensional data spaces, and scales well in case of growing files.

As our project goal is to bring a contribution to E-learning domain, our idea is to structure didactic chapters as concept maps and provide an efficient electronic students distribution by their results. The chapter's notions will be split depending on their level of priority as follows:

- Low priority notions – referring to introductive notions about the chapter.
- Medium priority notions – referring to basic notions of the chapter.
- Maximum priority notions – referring to advanced notions of the chapter.

As our concept maps are represented as tree structures, where each path of the tree is assigned a weight, these priorities can be computed depending on the assigned weights.

Assuming that, in order to evaluate a number of students, each chapter presents a final quiz, we have decided to use concept maps as weighted trees in order to generate tests containing notions of different priority levels. The algorithms responsible for generating tests are based on tree searches methods.

3 Building Clusters with M Tree and k-Means Algorithms

Our implementation uses the following student's data structures:

```
struct Student{
    int IDStudent;
    float Lp;
    float Mp;
    float Maxp;
    int Where[3];
};
```

The *IDStudent* is the identifier corresponding to each student entering the online distribution program. In order to evaluate these students, each chapter presents a final quiz containing notions belonging to the previously discussed levels of priority. Thereby , each student will

get a mark for each type of notion, contained in the chapter:
- a *Lp* mark corresponding to *low priority notions,*
- a *Mp* mark corresponding to *medium priority notions,*
- *MaxP* mar corresponding to *maximum priority notions*.

Each one of these marks is important , because they represent the guiding tool for a student. Example:

Let us suppose that the student identified by his/her *IDStudent=1002* takes the quiz, at the end of the chapter and gets the following results: (Lp= 7.70, Mp=6.78, Maxp=5.00).

Right know, a teacher, or even an electronic program, is able to compute the minimum performance of this student, reaching the conclusion that the advanced notions of the chapter(indicated by Maxp) have not been covered properly by this student, and therefore , he/she needs to put more energy in this direction. These directions are given by *Where* vector, used for providing instructions regarding where should a student improve his/her level of knowledge. A graphical representation is given below:

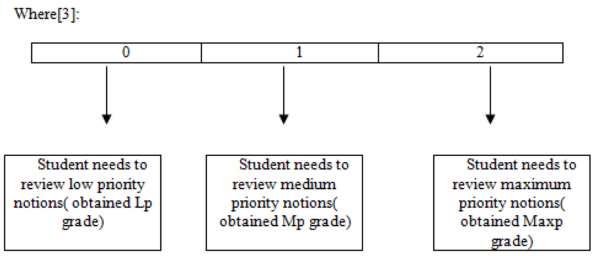

Figure 1: *Where* vector, used for providing instructions regarding where should a student improve his/her level of knowledge.

As a start, Where[3]=(0,0,0). If any of this vector's component changes to 1 , this means it becomes active. For instance, if StudentA.Where[3]=(0,1,0), it means that he/she needs to review notions of medium priority level.

These students are then, distributed and placed by our algorithm in a M-tree structure. Before moving on with the algorithm, let us present the structure of the M-tree. As it was explained above, the M-tree is a spatial, metric tree, consisting of: *1 root* and *k leaves* containing students.(As we will see later, these leaves represent clusters of students). For now, let us stick to the structure of a M-tree, mentioning that this tree is actually a spatial one, where its leaves can be imagined as spheres, containing points, which are actually students. As far as the structure of a node is concerned, we have:

The M-Tree node's structure is:

```
struct m_Node{
    int nrKeys;
    bool isLeaf;
    float radius [NMAX];
    m_Node *routes [NMAX];
    struct Student students[NMAX];
};
```

The `nrKeys` represents the number of students contained in a node (cluster). As far as our M-tree is concerned, we pay extra attetion to the nodes, because it is very important wheather they are leaves(terminal nodes) or internal nodes, as we will see later in the algorithm. The boolean variable `isLeaf` is pointing out our exact concern: weather a node is leaf or internal. Moving on, as itwas previously said, we consider our nodes as spheres, and as any sphere, it is geometrically represented by its center *C(x,y,z)* and its radius *R*. However, in order to match these notions to our real implementation, we have constructed an abstract interpretation for this geometrical representation. The center *C* of a sphere(cluster) will be represented by the average student belonging to the set of students contained in that particular cluster. Instead of spatial coordinates *(x,y,z)* , our *centerStudent* will be represented by its elements *(IDStudent,Lp,Mp,Maxp)*, which were presented earlier. The *radius* of this abstract sphere will be represented by the distance between the *centerStudent* and the student (students) with the lowest results in that cluster. We will take a closer look to this abstract system later, when we will discuss the implemented algorithm. The *routes* represent the children of a particular node and of course, the `nrKeys` points inside the sphere, which are actually the *students*, as in our abstract system, a spatial point is represented by a student.

Our implementation is based on the idea of students distribution depending on their results to a quiz at the end of a chapter. For a better understanding of our implementatio let us consider a real situtation:

Let us consider a finite set S of k students defined as S={St_1,St_2,….St_k), k>0. Let us supose that all these students have taken a quiz at the end of a chapter in order to evaluate their level of knowledge. Each student is represented by his/her IDStudent, and his/her grades: Lp, Mp, Maxp (they were discussed earlier). Let us assume that we want to create an hierarchy among these students, depending on their results. In order to do that, we need to group these students in *clusters*, each cluster having its own attribute. An attribute, for a cluster, represents the level of performance for that particular group of students. Moreover, these attributes are also used as indicators pointing out to the type of notion (low priority, medium priority, maximum priority) the student needs to review. After a group of students (cluster) is formed, a center is chosen, that is the average student in that group, and all the other students are distributed in a spherical manner, arround him.

Computation of a radius for a cluster : the radius of a cluster represents the maximum possible distance between the *centerStudent* and the rest. The bigger the distance is, the better or the lower the results of that student are. Just as in real cases, when we say there is a big distance between this *average student* and *student A*, this means that Student A has either better results or worse results, we don't know for sure. Anyhow, should the distance between the *centerStudent* and any other student be greater than the cluster's radius, it means that the particular student does not belong to that cluster, for the simple reason that he/she is smarter than all those students in that cluster or his/her results are lower than any other's in that cluster.

The radius of a cluster is computed, depending on the results of each student, being the maximum distance between two students. We define the distance between *StudentA and StudentB* as follows:

$$d_{AB}= \max\{(|Lp_A-Lp_B|, |Mp_A-Mp_B|, |Maxp_A-Maxp_B|)\} \quad (1)$$

As you can see, when we measure the distance between two students, we are looking for the most marking difference between them. This also helps us in defining attributes of a cluster, depending on the type of notion students should focus on (low priority notions, medium priority notions, maximum priority notions). Moreover the relation (1) guarantees that the radius of a cluster represents the biggest difference between the levels of knowledge for each student belonging to that cluster.

As example, let us consider the student A with his/her results: (9.60, 8, and 7.50). Let us consider the student B with his/her results: (7.60, 8, 6.50). Following the relation (1), we get the biggest difference 2 (9.60-7.60). This is the biggest distance between them two. So they might have similar knowledge for medium priority notions (8,8), and maximum priority notions(7.50,6.50), but when it comes to low priority notions, we see a gap between them(StudentA -9.60 , StudentB -7.60). Let us consider that StudentB has the lowest result in the cluster, and StudentA is the centerStudent. Then, as we have presented earlier , they can be grouped in a cluster with its radix 2 . Let us suppose now that StudentC gets the results: (5, 6.30, 5). We get:

$d_{AC}= \max \{|9.60-5|,|8-6.30|,| 7.50-5|\}=4.60$,

$d_{BC}= \max \{|7.60-5|,|8-6.30|,| 6.50-5|\}=2.60$.

As you can see, neighther of these distances is lower than our cluster's supposed radius, as the difference between Student and the students StudentA, StudentB is huge, so there is no way, StudentC will not become a member of this cluster. Moreover, based on the present results of StudentA and StudentB , we can define an attribute for this cluster: all students belonging to this cluster, will posses similar knowledge levels for medium and maximum priority notions, but the marking difference between them will be represented by the low priority notions, so all of them need to review this part of the chapter.

The main steps of the agorithm are:

Step1. We start from a simple representation of students, identified by their elements:

- IDStudent
- Lp (score)
- Mp (score)
- Maxp (score)

Step2. We picture the set of the students who have taken the test as the points in 3D space. Our algorithm involves two major operations:

- a clustering operation
- a split operation

We will first describe the *spliting* method. We have decided that these groups of students should have a maximum number of allowed members. Let us denote this number as the *filling factor* of a cluster (student group). Whenever the number of students in a particular cluster becomes greater that this filling factor, a cluster splitting is involved. This is how the M-tree extends its nodes. The splitting procedure works as follows:

At the beginning two random students from that cluster are chosen as the centers for the new clusters resulting after splitting. Let us denote them *Student1* and *Student2*. Next, we distribute the rest of the students arround the new centers *Student1* and *Student2*. If, for instance, we have *Student1* and *Student2* as centers, the question is where should we attach *Student3* to? We compute the distance between (Student1, Student3) and *(Student2, Student3)*, using relation (1). *Student3* will go near that student which is more closed to him/her (that is, from a level of knowledge point of view). After the distribution is completed, we start the *chooseCenter* method, which recalculates the new centers of the clusters, and if new centers are found, the entire discussed process happens again, until new centers are found no more. This process is called the *Clustering Process*. After that the effective splitting happens, and the initial tree node is split into two nodes. When these clusters are formed, they are also assigned atributes(we have discussed about them earlier). What is interesting is that these atrributes suffer a constant evolution, depending of the students inserted in that specific cluster. As the clusters suffer constant splittings and modifications, whenever the number of students inside is greater than the filling factor, so do the attributes change, transforming a part of the old cluster in a better one (shelters students with higher levels of knowledge) or even a lower one (shelters students with lower levels of knowledge).

The classical M Tree algorithm has been adapted such that the final structure has two levels. The procedure for building the structure takes into consideration both the desired number of clusters and the filling factor of a leaf node.

```
procedure MTree (x1, x2, …, xN; K; F)
// K – the number of clusters
// F– filling factor
 for ( i=1, i<N ){
   Ci = FindCentroid (centroids, xi);
   if ( #Leaf [Ci] has F instances)
     if (#we have k clusters)
       #put xi in Leaf [Ci]
     else
       #split Leaf [Ci]
   else
     #put xi in Leaf[Ci]
   RecomputeCentroids(Leaf[Ci])
 }//end for
```

The computational complexity of this M Tree procedure takes into consideration that each instance is considered only once. That is why the complexity of this operation is *O(N)*. Still, the number of clusters influences the complexity since the best corresponding cluster needs to be determined. The time taken for this operation is *O(K)*. Still, the recomputation of the centroid is not so costly as in the case of k-means algorithm due to the filling factor parameter. Thus, the overall complexity of employed M Tree procedure is O(KNF).

Clustering and splitting procedures in pseudocode are presented below.

```
Procedure Cluster_and_split
  #Get the cluster which will suffer splitting procedure;
  #Let StudentSet={Si| Si є cluster, i=0, nrKeys, Si student};
  #Choose
        S1 є StudentSet (as new center);
        S2 є StudentSet (as new center2);
        oldCenterStudent1←S1;
        oldCenterStudent2←S2;
  #While (!STOP_CONDITION) do {
        For each other Si є StudentSet, (Si ≠ S1 and Si≠ S2 ) do {
     Compute d1i=max{(|Lp1-Lpi|,|Mp1-Mpi|,|Maxp1-Maxpi|);
     Compute d2i=max{(|Lp2-Lpi|,|Mp2-Mpi|,|Maxp2-Maxpi|);
     If (d1i < d2i)
        Attach Si to the cluster with center S 1
     else
        Attach Si to the cluster with center S 2
     #Determine attributes
    }//end do

newCenterStudent1=chooseCenter(newFormedCluster1, newClusterRadix1);

newCenterStudent2=chooseCenter(newFormedCluster2, newClusterRadius2);
   STOP_CONDITION←
        (newCenterStudent1=oldCenterStudent1)
        && (newCenterStudent1=oldCenterStudent1))
  #Effective_split_of_initial_cluster
  }//end while
```

The *Determine attributes* sequence works in the following way. Every time a distance between two students is computed, after the marking difference between them is extracted, inside the *Where[3]* vector, the corresponding component of that type of level notion the maximum was extracted for, is incremented (we say a flag is raised for that component).

Example. Let us consider that the marjing difference between studentA and studentB regards the low priority notions, then studentA.Where[0]=1 and studentB.Where[0]=1 (see the *Where[3]* vector configuration in previous sections). Notice that this *Where[3]* vector changes for every new student in the cluster, because in the end of the clustering process only the attributes of the average student will prevail, because he/she is the center of that particular cluster.

The *chooseCenter* sequence simply computes the biggest distance between all the students and also sets the new studentCenter of the cluster, as we will see in the complete algorithm's pseudocode.

Now we will take a closer look to the *splitting procedure*:

```
Procedure Effective_split_of_initial_cluster
    m_Node = node for splitting
  m_node→isLeaf = false;
  //the initial node becomes a root,
  //where the centers of the clusters will be retained
  # Alloc memory for leftChild and rightChild of m_node;
  m_Node→leftChild = newCluster1;
  m_Node→rightChild = newCluster2;
  #Insert S = { centerStudent1|radiusCluster1,
        centerStudent2|radiusCluster2} in m_Node
  #Attach the rest of students Si
        to leftChild or rightChild
  leftChild→isLeaf=true;
  rightChild→isLeaf=true;
```

After seeing these two important steps in the algortihm, it is time to present the entire process:

```
Procedure Build_M_Tree
   #initializeTree;//create an empty root tree
                              //and set root as leaf
   #While (not endOfFile){
     #Get input data student(IDStudent, Lp,Mp,Maxp);
     #If   ( tree->root->nrKeys < filling factor )
       #make a simple insertion in the actual root
     #Else
       #apply Cluster_and_split;
   }//end while
```

The classical standard k-means algorithm partitions a dataset on N instances into K clusters. The algorithm is:

```
procedure k-means (x1, x2, …, xN; K)
  {c1, c2, …, cK} ← Select Random Centroids
  for ( k=1, k<K )
    centroidk = ck;//these are initial centroids
  while (#centroids are not same){
   for ( k=1, k<K ){
     for ( n=1, n<N ){
       j = index of corresponding cluster
       #put xn in corresponding cluster Cj
     }//end for
   }//end for
   for (k=1, k<K)
     # compute centroids for all clusters
  }//end while
```

The most important discussion regards the computational complexity of k-means algorithms. Most of the time is spent on computing distances between items. This computing is performed when putting instance x_n in cluster C_j. One such operation costs *log(M)*. The reassignment step computes *KN* distances, so its overall complexity is *log(KNM)*. In the re-computation step, each vector gets added to a centroid once, so the complexity of this step is *log(NM)*. For a fixed number of iterations I, the overall complexity is therefore *log(IKNM)*.Thus, K-means is linear in all relevant factors: iterations, number of clusters, number of vectors, and dimensionality of the space. One of the most important issues regards the number of iterations. In most cases, K-means quickly reaches either complete convergence or a clustering that is close to convergence. In the latter case, a few items would switch membership if further iterations are computed. This computation has a small effect on the overall quality of the clustering.

4 Clustering Evaluation Metrics

A comparison of two distinct procedures defines the needed steps in order to obtain sound results. The presented analysis procedure compares two methods that are used for building clusters of items: k-means and M Tree.

An input dataset is considered. Both k-means and M Tree algorithms are than used for building clusters from the same dataset. For each set of clusters, there are computed specific indicators for characterizing obtained clusters. The indicators that are taken into considerations are:

Tightness Indicator:

$$Q = \sum_{i=1}^{k} \frac{1}{|C_i|} \sum_{x \in C_i} d(x, \mu_i)$$

where $|C_i|$ is the number of points from cluster i. The value for Q will be small if the data points from the cluster are close. Thus, in the comparison analysis procedure the clusters with smaller computed values of Q have higher quality.

Homogeneity Indicator:

If the centroids of clusters are computed with formula $r_k = \frac{1}{n_k} \sum_{x \in C_k} x$, where x are the instances from cluster C_k , than homogeneity indicator is

$$H(C) = \sum_{k=1}^{K} \sum_{x \in C_k} d(x, r_k)^2$$

The value for H will be small if a cluster has homogeneous structure. This, in the comparison analysis procedure the clusters with smaller computed values of H have higher quality.

Cluster Distance:

$$CD = \sum_{1 \le j \le k \le K} d(r_j, r_k)^2$$

where j and k are indexes of clusters whose centroids r are taken into consideration. The value for CD will be big if the similarity among clusters themselves is low. Thus, in the comparison analysis procedure the methods with bigger computed values of CD have higher quality.

Weakest Link between Points:

The weakest link for a cluster is the maximal value of all pairs of points belonging to the same cluster.

$$WL = \max(d(x_i, x_j)),$$

for all x_i and x_j belonging to the same cluster.

5 Experimental Results

The first experiment builds an M Tree from the data from 6 students. The values for (Lp, Mp, Maxp) = {(8,7,6), (5,9,10), (7,7,7), (7,8,7), (6,6,6), (4,3,5)}. The value of filling factor is 4.

We insert students with IDStudent 1,2,3,4 in one cluster. Student 5 also needs distribution, but adding him/her to the same cluster is not possible, since a cluster can hold a maximum of 4 persons (filling factor is 4). So a split is mandatory.

After split the clusters are:

Cluster1={Student1, Student3, Student4},
Cluster2={Student2}.

Then, student 5 may be inserted, and the clusters become as follows:

Cluster1={Student1, Student3, Student4, Student5} ,
Cluster2={Student2}.

A new split is necessary, because cluster 1 is full. Thus, the following clusters are being obtained:

Cluster1={Student3, Student4, Student5}
Cluster2={Student2}
Cluster3={Student1}

Finaly, Sudent6 is inserted in cluster 3, and thus the following clusters are obtained:

Cluster1={Student3, Student4, Student5}
Cluster2={Student2}
Cluster3={Student1, Student6}.

The *chooseCenter* sequence simply computes the biggest distance between all the students, also it sets the new studentCenter of the cluster, as we will see in the complete algorithm's pseudocode.

The second experiment takes into consideration 15 students. For each student there are available two weighted parameters: the number loggings and the time spent on-line. The real parameters scale such that all values are in the range 0 to 16.

The input dataset is:

A1(10.94 , 11.86); A6(11.02,2.28); A11(9.29 , 13.86); A2(1.58 , 6.27); A7(11.23,9.37); A12(8.00 , 1.09); A3(13.66 , 4.62);

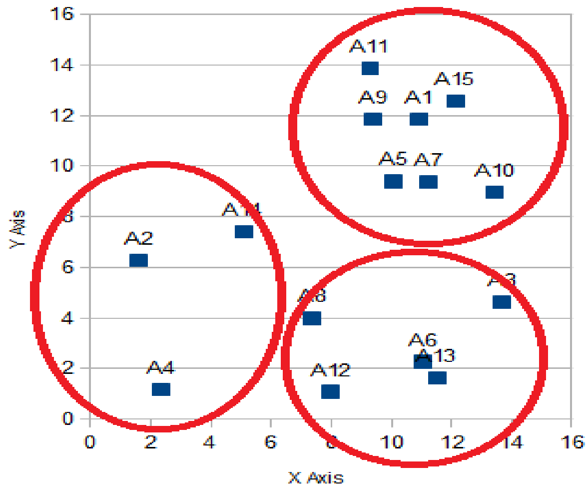

Figure 2: Distribution of learners with k-Means.

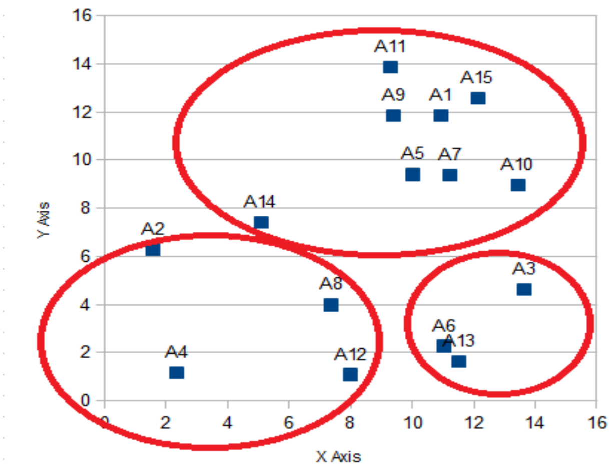

Figure 3: Distribution of the learners with MTree.

A8(7.35 , 3.99); A13(11.52 , 1.63); A4(2.33 , 1.16); A9(9.4 , 11.84); A14(5.08 , 7.42); A5(10.04 , 9.41); A10(13.43 , 8.97); A15(12.12 , 12.59);

This dataset is used for building three clusters with both k-means and M Tree algorithms. The obtained distributions are presented in figures 2 and 3.

The obtained clusters by k-Means clustering have the following centroids and composition:

C1 (2.99,4.95) *//Cluster 1's Centroid*
A2 (1.58, 6.27)
A4 (2.33, 1.16)
A14 (5.08, 7.42)

C2 (10.92, 11.13) *//Cluster 2's Centroid*
A1 (10.94, 11.86)
A2 (10.04, 9.41)
A7 (11.23, 9.37)
A9 (9.40, 11.84)
A10 (13.43, 8.97)
A11 (9.29, 13.86)
A15 (12.12, 12.59)

C3 (10.31, 2.72) – Cluster 3's Centroid
A3 (13.66, 4.62)
A6 (11.02, 2.28)
A8 (7.35, 3.99)
A12 (8.00, 1.09)
A13 (11.52, 1.63)

The obtained clusters by M Tree clustering have the following centroids and composition:

C1 (10.04, 9.41) - *//Cluster 1's Centroid*
A1 (10.94, 11.86)
A2 (10.04, 9.41)
A7 (11.23, 9.37)
A9 (9.40, 11.84)
A10 (13.43, 8.97)
A11 (9.29, 13.86)
A14 (5.08, 7.42)
A15 (12.12, 12.59)

C2 (2.33, 1.16) *//Cluster 2's Centroid*
A2 (1.58, 6.27)
A4 (2.33, 1.16)
A8 (7.35, 3.99)
A12 (8.00, 1.09)

C3 (11.02, 2.28) *//Cluster 3's Centroid*
A3 (13.66, 4.62)
A6 (11.02, 2.28)
A13 (11.52, 1.63)

For each clustering procedure there were computed the evaluation metrics presented in section 3. The results are presented in the following table:

Indicator	Clustering Procedure	
	k-means	M Tree
Tightness	7.55	8.52
Homogeneity	100.47	137.48
Clusters Distance	230.47	203.11

Table 1: Tightness, homogeneity and cluster distance indicators for k-means and MTree distributions.

The link analysis for both distributions is presented in the following table:

Indicator	Clustering Procedure	
	k-means	M Tree
Weakest Link Cluster 1	0.9	1.21
Weakest Link Cluster 2	0.84	1.15
Weakest Link Cluster 3	0.87	0.51

Table 2: Weakest link values obtained for k-means and MTree distributions.

The k-means results are obtained using Weka [4]. Weka is a collection of machine learning algorithms for data mining tasks which has implemented the k-means clustering algorithm.

The M Tree results are obtained, using a custom Java implementation of the algorithm. The main differences of this implementation compared with classical M Tree algorithm regard two aspects. One regards the general structure of the tree that is restricted to two levels. This means there is only one root node where centroids along with covered radius are placed. The second issue regards the way *k* (the number of clusters) and f (the filling factor) are managed. If the algorithm is required to produce a certain number of clusters, the instances are placed into appropriate clusters until a filling factor is reached. When this happens, a split is performed. Splitting is no longer performed when the desired number of clusters is reached. In this way, the clustering process is directly managed by the values *k* and *s*.

The comparison of the two obtained distributions reveals the fact that the M Tree distribution clusters have lower quality than the ones obtained by usage of k-means. Still, the results obtained by M Tree are very different from the ones obtained by k-means. All indicators presented in table 1 have better results for k-means than the ones obtained for M Tree. It can be observed that the tightness and homogeneity are better (because they have smaller values) for k-means than for M Tree.

Another comparison that may be done regards the mobility of centroids. Although the differences of computed indicator (Tightness, Homogeneity, Clusters Distance) values are not very small, the computed centroids are quite close.

6 Conclusions

The paper presents a study usage of an implementation of M-Trees building algorithm. The tree manages real data representing e-Students (students from an e-Learning environment). The instances (i.e. the students) are characterized by attributes representing the scores obtained when taking tests. The tests are classified according with their level of difficulty: low, medium and high.

Within classical M-Tree building procedure it was used a custom clustering procedure when splitting a node was necessary. The clustering procedure is designed such that produces an optimal grouping of students regarding the "distances" in knowledge among them.

The tests were performed with datasets representing 200 students, and the filling factor of a cluster was restricted to 18. As a result, we got 7 clusters, with attributes, leading e-Students to notion reviews.

Another goal of our current implementation is to provide valid data distribution using specific data validation algorithms. Moreover our algorithm is able, for the moment, to distribute e-Students which test their level of knowledge for one chapter only. We wish to extend this process for multiple chapters.

This paper also presents a procedure that measures the degree in which the effectiveness of an e-learning process has improved. The analysis process is data centred. The data represents experiences provided by learners. In this study two features (attributes) characterize each learner: the number loggings and the time spent on-line.

The goal of the procedure is to produce clusters of users using two different techniques: standard k-means algorithm implemented in weka and a custom flavour of M Tree algorithm with a custom implementation.

The input dataset is restricted to a sample of 15 learners. This choice is because a manual inspection of the obtained clusters is desired. An automated analysis of the obtained clusters is performed by computing some basic clustering quality metrics: Tightness, Homogeneity, Clusters Distance and link analysis. The obtained results show an acceptable quality of the M Tree clusters although the computational complexity of the algorithm is much lower than complexity of k-means.

The main goal of the paper is to find an algorithm that produces acceptable results with complexity much smaller than a classical procedure.

The quality of the obtained clusters has a direct influence over the degree in which the e-learning process has been performed. Unsupervised classification (clustering) is one of the main methods for making evidence regarding the knowledge acquisition of learners. Once a high quality distribution has been discovered, a learner may by clustered at certain moments and progress may be evaluated. Of course, the process needs to be well defined and needs to be based on a high quality clustering procedure.

The future works regard different aspects. A first issue would be to replicate the procedure with more data. This may be accomplished on hundreds or even thousands of learners, if data are available. The clustering procedure is highly influenced by the initial centroids. In custom initialization is advisable. A good starting point may be obtained by using a k-means clustering on a sample dataset from the entire dataset. The quality of the clustering process is directly influenced by the choices made regarding k and f values. Thus, an initialization step may also refer to prior computation of the optimal number of clusters and optimal filling factor. The computation of these parameters may be delegated to other high quality clustering procedure that works on a data sample.

Finally, there may be defined procedures for assessing progress in time and even recommendations. The progress in time may be computed classifying the learner from time to time. This may yield to a learning path that has been followed by the learner. More than this, there may be obtained recommendations for the learner. The recommendations may regard necessary actions necessary to be taken by the learner in order to improve his learning curve.

References

[1] J. Hartigan and M. Wong (1979). A k-means clustering algorithm. In *Applied Statistics*, 28, pp. 100–108.

[2] Uhlmann, J.K. (1991). Satisfying General Proximity/Similarity Queries with Metric Trees. *Information Processing Letters*, Vol. 40, pp. 175-179.

[3] Ciaccia, P., Patella, M., Zezula, P. (1997). M-tree: An Efficient Access Method for Similarity Search in Metric Spaces. In: *VLDB'97, Proceedings of 23rd International Conference on Very Large Data Bases*. Morgan Kaufmann, pp. 426-435.

[4] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009). The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, Vol. 11, Issue 1.

[5] Burlea Schiopoiu A. (2008). The Complexity of an e-Learning System: A Paradigm for the Human Factor, *The Inter-Networked World: ISD Theory, Practice and Education*. Vol. 2, Springer-Verlag, New York, pp. 267-278.

[6] V. Fomichov and O. Fomichova (2006. . Cognitonics as a New Science and Its Significance for Informatics and Information Society. Special Issue on Developing Creativity and Broad Mental Outlook in the Information Society (Guest Editor Vladimir Fomichov), *Informatica. (Slovenia),* 2006, 30 (4), pp. 387-398.

[7] C. Romero and S. Ventura (2007). Educational Data Mining: A Survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), pp. 135-146.

[8] Gema Bello Orgaz, Héctor D. Menéndez, David Camacho (2011). Using the Clustering Coefficient to Guide a Genetic-Based Communities Finding Algorithm, *IDEAL 2011,* pp.160-169.

Computer-Aided Educational Intervention in Teenagers Through Internet Social Networking

Maria Guadalupe Velázquez-Guzman
Universidad Pedagógica Nacional, México
E-mail: gvelazg@gmail.com

Felipe Lara-Rosano
Universidad Nacional Autónoma de México, México

Keywords: Internet social networks, educational intervention, school practices

Received: December 15, 2011

Educational institutions have the great challenge and the commitment not only to teach mathematics, science and language. They also must contribute to forming the new citizens. These new citizens should be critical and creative social actors that participate in the construction of better ways of coexistence. In this paper we propose educational strategies based on the use of Internet social networking sites as Facebook®, Twitter® and MySpace®. These strategies must enhance the social development of the students, taking into account their peers' subculture.

Povzetek: Predlagane so učne strategije na osnovi socialnih omrežij.

1 Introduction

The school, as educational institution, is the essential nucleus for the formation of future citizens and social actors [1]. In fact, it is the place of meeting of diverse individualities and social activities where values are affirmed, as well as forms of being and coexisting.

However in our contemporary societies, from the least developed to the so called "first world" countries, we find now a social environment of great violence and brutality that influences in a negative way the development of the young people, driving them toward an aggressive cruel behavior.

Therefore the educational institutions have the great challenge and the commitment not only to teach mathematics, science and language. They also must contribute to the reconstruction of the social fabric by forming the new citizens. These new citizens should be critical and creative social actors that participate in the construction of better ways of coexistence. According to Stacey [2], when people communicate with each other to accomplish the joint action of living and acting together, they are continuously relating to each other in a responsive manner. Marion [3] asks what catalyses social behavior and concludes that humans are said to cooperate because that is the best way of achieving individual goals. In addition, individual human action is catalyzed by symbols: ideas, concepts, opinions, beliefs, emotions, projections and values like the understanding of the personal differences, the cooperation for the well of the collective whole, the search of justness and equity and the practice of tolerance.

In this paper we identify in the actual school practice obstacles and facilitators for the individual and social development of the students for living together and propose educational strategies based on computer-aided Internet social networking such as Facebook® and Twitter®.These strategies must enhance the social development of the students, taking into account their peers' subculture.

2 The School Social Development Objectives for Living Together

Let us consider a school group community as a complex system composed by the teacher and the students peer group, defined as a collective of human agents that interact with one another in a common internal environment defined by the school rules and the group interactions. According to Mead [4], it is in the detailed interaction between people, their ongoing choices and actions in their relating to each other that their minds and selves arise. However, the schools are located in neighborhoods where are seated families that belong to very diverse socioeconomic strata. In the school the adolescents meet having different socialization processes. Muuss [5] affirms that the socialization of adolescents depends on the family and the neighborhood where the individuals live and develop.

Therefore in the school there is a great diversity of individuals with different behavior patterns, guided by a multiplicity of beliefs, values and customs that give place to misperceptions frequently in conflict. The socio-cultural differences of individuals and families present a diversity of values, beliefs, lifestyles and behavior rules that are a part of a plural society. This diversity is reflected in the students peer group. In fact, these diverse individualities, nevertheless the differences, have a strong tendency to constitute peers' groups where adults are left out. According to Stacey [2] "organizing is human experience as the living present, that is, continual interaction between humans who are all forming intentions, choosing and acting in relation to each other as they go about their daily work together". Handel [6] affirms that the peers' groups are complex organizations that perform a socialization function among their members. In these organizations the members establish goals and make agreements directed to certain matters, interests and concerns. These agreements must follow certain procedures.

Savin[7] sustains that from a perspective of the human psychological development, it is important that inside a net of interactions the individual finds a place for itself, a status. In this net it is necessarily to have a pattern of power relationships based on the different capacities and personal abilities, being this pattern a constant feature in the structure of human relationships.

In fact, the adolescents have such a great necessity to belong to a peers' group that they agree to accept the conditions that they must cover in order to be accepted. These conditions depend on the group's identity, its own values, belief and behavior systems and are related to the image that the group wants to express. Among these requirements could be mentioned the physical characteristics, certain specific abilities, and a sense of humor toward himself.

The peers derive their standards to think and to judge, their individual references, behaviors and values from their group. According to Corsaro [8], many activities of peers' interaction look for a sense of self and opposition to the adult world. The peers' group defines types of relationship that are qualitatively different from the relationships with the figures of authority. Some of these new types of relationship are reciprocity (not possible in the relationships with the adults), identity, belonging, solidarity and support. The peers feel that they participate at the same level in decision-making and the establishment of norms for the group. Certainly, the group rewards its members, but it also punishes them with distancing, ostracism and other disapproval expressions.

In accordance with this teenage subculture, the objective of forming citizens that have the sensibility of participating with the other ones in the construction of a living together community, it is a work that is one of the main challenges of the education. In fact, the school must form citizens to be able:

a) To understand and accept the principles and forms to live together in a heterogeneous community.
b) To understand the common benefits of collaboration and mutual help.
c) To balance the respect to the individuality with the collective good.
d) To recognize and respect the personal differences.

Evaluating in the actual school practice the accomplishment of these objectives, we found [9] the following conflicts in the practice implementation that hinder the educational targets:

- The prevalence in the school of the control, surveillance and punishment measures. McGregor [10] distinguishes between autocratic and participative styles of group management. He argues that autocratic styles are based on the view that people dislike work and avoid it if they could. They need to be coerced into working, being punished if they do not work and rewarded if they do. This is the "stick and the carrot" approach that is still so prevalent in schools.
- The absence of dialogue between teachers and students. It is important that the rules are the result of a practical ethical knowledge. Therefore, they should contemplate the rights of everybody. [11].

The previous results let us pose the following questions:

- Which educational strategies could be implemented in the school for developing effectively the future citizens?
- Is it possible a different organization from the current conventional forms?
- Is it possible that the teachers as educational agents go so far as to modify their behaviors in order to become better builders of future citizens?

3 Proposals for a New Educational Intervention Paradigm

It is important to recognize that the adolescents possess natural organizational powers that it is important to guide with the educational intervention for their own social development. Each individual is like an attractor, and when individuals come together to form a group, they resonate with each other, producing through their communication a social attractor [9], [3].

The word potency refers to the characteristic quality of the human being to advance toward objectives when the person is endowed with self-reliance and is able to accomplish her objective [3]. However, potency is not enough. The person needs a reference frame about the natural and social world, that should be structured and have an internal coherence, where she should find her place. Without this framework she will get confused and will be unable to act upon in a proper way. Savin's reflection is similar [14], since he refers to the importance that the individual finds a suitable place inside a net of interactions.

Moreover, the adolescent also needs a set of objectives as the focal point of her actions and the expression of her effective values. This necessity to have defined objectives responds to an existential primary need that demands satisfaction, without caring the means.

The contributions of Fromm [12] and Savin[7] are the base to understand the adolescents' auto-organization, the individual necessity to belong to a peers' group and the origin of the leadership that some of them impose over this organization. However when these necessities are ignored then the adolescents in their social diversity are not properly guided. They end up forming contesting groups mainly with tendencies of untamed dominating power.

Sherif and Sherif[13] suggest that the individuals must be guided toward humanist and not destructive goals, because their nature is malleable, notwithstanding that by nature they differ in capacity, abilities, resources and experiences. It is starting from this differentiation that they can contribute to the solution of their problems. The uniformity is not a source of strength in a group.

What we want to emphasize is that in the adolescent peers' groups we can find the organizational potency, the search of goals, the values of reciprocity, solidarity and belonging. However, without an educational guide from the school this emerging social organization can take any form.

Taking into account that the maintenance of a relationship is dependent upon the members each striving to reach an acceptable balance between their own desires and needs and that of others, according to the approach of Relational Dialectics [14], we developed the following proposals [9]:

1) That the dialogue and the agreements be the means to solve novel situations. Following the perspective of Gadamer [15] about the community sense, the practical knowledge that arises through the contrasting experience is a phrónesis, because it is knowledge to distinguish what is good and what is bad for the community. This moral knowledge subsumes the individual preferences to the communality correct, making familiar to the students a community sense that before was strange. These dialogue and agreements may be easily implemented through Internet mediated social networks discussions through the creation of communities of practice [16], [17], [18] in the classroom with communications through Internet resources like Facebook®. According to Miranda-Pinto and Osorio [19], in a community of practice supported by the information and communication technologies the virtual interactions define new ways to reinforce the cooperation and the knowledge construction among peers.
2) That teachers and pupils, by means of the dialogue and the argumentation, participate in the making and upgrading of the rules, defining collectively the principles, rights and responsibilities of the members of the community to make possible the coexistence for the common good, nevertheless the plurality and the personal differences. In fact, 59% of students with access to the Internet [20] report that they use social networking sites to discuss educational topics including career and college planning, and 50% use the sites to talk about school assignments. Some parents and teachers say that using these sites helps students improve their reading, writing, and conflict resolution skills and learn to express themselves more clearly.
3) That the values that strengthen the community sense like friendship, respect, participation, solidarity, common good and the feeling of belonging to a community are encouraged. Under the surface of the social, individual resonances harmonize in the sense that people develop a shared view. Prigogine [21] casts his theories in terms of entities resonating with each other and evolving as collective ensembles. Therefore a strategic policy enhancing the above values through the use of Internet based social networks may result in the development of social attractors that reproduce themselves changing the social pattern.
4) That the rules take into account the diverse needs for the members and their interactions. Langton [22] talks about the inability to provide a global rule, for changes in a complex system's global state, making it necessary to concentrate on the interactions occurring at a local level between agents in the system, like those of social networks. He states that it is the logical structure of the interactions, rather than the properties of the agents themselves, which is important, thus potentially elevating interaction to primacy.

These proposals are further supported taking into account the organizational power of adolescent peers' groups.

4 Organizational Power of Adolescent Peers' Groups

In the school the adolescents, in spite of their differences, have a strong tendency to constitute peers' groups where the adults are left out.

The Internet social networking as Facebook®, Twitter® and MySpace® and photos and video sharing sites such as YouTube® have emerged as new media very popular among adolescents and have shown a big impact on the organization of adolescents peers' groups.

We think that this trend can be developed educationally, putting emphasis in its positive potential [12]

The school has a functional structure that could be used to guide the suggested peers' group policy. The school has teaching and learning objectives, collective projects and practices, implicit socialization actions in the teacher-student relationships, behavior rules and sanctions. All these organizational elements should be used to implement an integral adolescent socialization program (See Figure 1).

In this program the adolescent internal values that are shown in the peers' culture like reciprocity, identity, support, self-reliance, friendship, belonging, solidarity, organization, recognition and respect should be reinforced and consolidated as behavioural attractors of

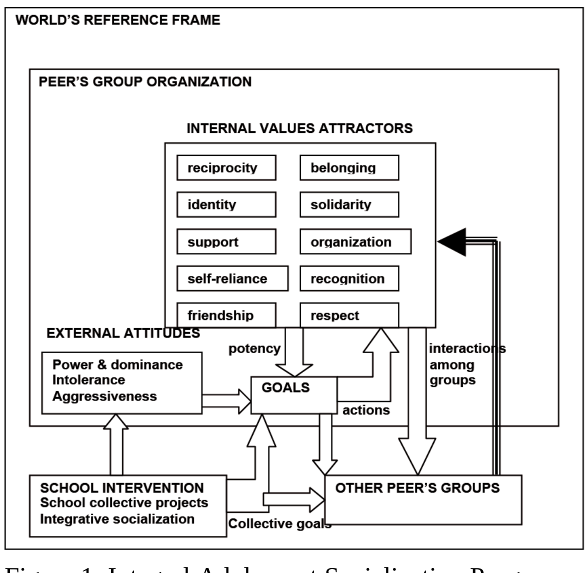

Figure 1: Integral Adolescent Socialization Program.

the adolescent's culture. This policy will permit to get the socialization goals to overcome the power and dominance attitude, the intolerance and aggressiveness not only among individuals but among different peers' groups.

Some school projects like the organization of school sport teams to compete against other schools, the participation in science fair projects, theater festivals, school musical bands and community projects imply the student's individual active participation in community activities. These community activities could be the means for the students to become responsible social actors able to contribute in organizational forms for the social coexistence and collective good [9]. Individuals join groups, motivated by internal values attractors, creating a collective whole that is more than the sum of their individualities. If some of these internal values are related with living together in a heterogeneous community with a balance between the respect to the individuality and the collective good and participating in the definition of the collective principles, rights and responsibilities of the members of the community then this peers' group will affect the way of how their members evolve [11] improving the coexistence with others peers' groups.

The challenge for parents and schools alike is to eliminate the negative uses of electronic media while preserving their significant contributions to education and social connection [20].

In fact, it is important that inside a net of interactions (like Facebook® and MySpace®) the individual finds a place for itself, a status [23]. In this net it is necessarily to have a pattern of power relationships based on the different capacities and personal abilities, this pattern being a constant feature in the structure of human relationships.

5 Organizational Elements for an Intervention Proposal in the School

The school has a functional structure that could be used to guide the suggested peers' group policy [9]. All these organizational elements should be used to implement an integral adolescent socialization program through the implementation of school communities of practice supported by Internet social networks sites like Facebook®, MySpace® and Twetter®. Examples of these collective projects and practices to implement an integrative socialization of the students through the social network media are the following:

1) The discussion of actual problems affecting directly the young people like VIH, drugs and other health issues, professional training, labor insertion, family relationships, etc.
2) The discussion of community level problems, like security problems, civil rights, civil protection plans for emergencies (earthquakes, floods, tornados, landslides, etc), community waste management, sustainable community development, etc.
3) The discussion of global world problems like global economic crises, sustainable development, global warming, global epidemics, global education, etc.
4) The individual contribution of the student to the work in class (presentations, discussions, questioning, inquiring).
5) The organization of educational games and sport teams.
6) The organization of the working environment in the classroom.

A community of practice supported by Information and Communication Technologies (ICT) used as a form of participation can promote different social experiences as well as a process of self-training. To participate in these communities is necessary to consider the following aspects.

1) Internet connection, personal computers and a proper platform where the community is structured.
2) The communication established in the community can be done in two ways. Synchronously through chat where communication occurs in real time with those who are online. Asynchronously, through dialogue forums that allow a more extended participation without being all in line. This last type of communication allows everyone to control their participation, re-read the previous messages and ask for any more information.
3) The sense of community develops when the participants are active elements trying to reach common goals.

6 A Methodology for the Operation of School Communities of Practice Supported by Internet

We propose below a participatory synthetic methodology based on group dynamics for the operation of school communities of practice supported by Facebook®. This methodology has been proved successfully in college level students and its application to secondary and preparatory school levels is under way.

The methodology consists of the following eight phases:

a) *Integration of the specific community of practice*. The community of practice involves the students of the school class participating in the socialization program. The students should register at Facebook® with an alias or nickname of their own choosing, in order to maintain anonymity as possible to allow them to break away from emotional factors, friendships and enmities during the exercise for greater objectivity of the procedure. The participant list is sorted alphabetically by surname. Then the teacher explains the background to the problem to be analyzed, why he/she has called students as experts to be consulted, the objectives to be achieved through consultation and the mechanics of the session, emphasizing the importance of collaboration of everybody.

b) *Individual consultation*. The group meets in a classroom. The teacher asks an appropriate question for purposes of consultation and leaves as a homework to each of the participants to give three to five responses that are relevant. It is desirable that the number of responses does not exceed seven to only include the most important ones. Sometimes it is convenient to give an indication of the expected number of ideas per participant, for example, "enter at least five main factors that affect X". Sometimes the participant should select only those ideas that are most important, in which case it is asked, for example, "write the five factors most important for X".

c) *Exchange.* Each participant submits through Facebook® each of its n responses to a different participant, beginning with the participant who is next in the alphabetical list of participants. This process is repeated until no participant has responses to distribute. Upon receiving the responses of others, everyone should read them in order to be familiar with its contents. In case of doubt, the author of the response in question must clarify it. In this process each author is detached from their own ideas, focusing on those of others, creating an additional degree of anonymity that promotes the objectivity of the discussions.

d) *Grouping*. Then the community of practice meets again in the classroom for a Facebook® forum. In turn, each participant send one of the responses he/she got to the group. This selected response is called stimulus response. If any of the other members of the group believes that there is some relationship between this stimulus response and any of the responses she has, then he/she takes turn to speak and, after writing and sending his/her response to all, asks for approval of the community(“like” or “not like”) to group this response with the stimulus response. The grouping criterion depends on the problem, and the teacher must clarify any doubt about it. Sometimes it is convenient to group similar ideas, or ideas related to the same subject, or phenomena that recognize the same cause or have the same consequences.

If he/she gets the approval of the majority, his/her response is put in a list together with the stimulus response. This process is repeated until no more responses can be related with the stimulus response. Then another participant selects and sends another response to the group as a new stimulus response, and the association process is repeated. This process goes on until all responses of all participants have been grouped in lists. Each list ist held by the owner of the corresponding stimulus response. In some cases, a list can have only one isolated response.

e) *Synthesis*. The holding participant analyzes the content of each of the lists that he/she has and proposes a synthesis of the content expressed in a sentence in a few words. The sentence should not be a simple aggregation of the content of the responses, but indicate the common essence of them. Likewise, each response must logically imply the synthesis. Finally, the synthesis should be as specific as possible and contain the maximum information.

f) *Dialectic*. In a new community Facebook®forum, the author of each synthesis will send it to the other participants, together with the corresponding responses. This synthesis will be the starting point for a group discussion until the group adopts a final synthesis, which will be written as the final title of the list.

g) *Iterations*. Once titled all the lists, they will be subjected to a new process of association, repeating the steps of grouping and dialectical synthesis (in one or several iterations) until there remains only three to seven large general statements containing a hierarchical structure of minor statements.

h) *Presentation of the results*. Then the results are expressed as a schematic index of contents, with the major general statements as chapter titles and the hierarchical subordinated statements as sections, subsections, etc. of these chapters.

This technique must be adapted to each case. The teacher's role will be to introduce in each case the modifications and extensions necessary to achieve the desired goals, including combining it with other group dynamics techniques.

This technique has the following advantages:

a) It stimulates individual creativity, forcing participants to generate ideas.
b) It promotes the concrete expression of ideas by requiring that they are properly written.
c) It avoids the spread in the discussions, focusing the discussion on very particular points.
d) It ensures logical, systematized and prioritized outcomes, resulting from agreements.

Furthermore, with respect to the participants, this technique allows the teacher to:

a) Involve the participants in a process of awareness and sensitivity in order to correct for self-discovery their wrong attitudes.
b) Integrate a team that worked before as incoherent group.
c) Increase and harmonize the relationships of each team member.

This technique through Facebook® is applicable both in the definition of what is desirable, as in the exploration of possible futures, plus it can also be used in the process of identifying objectives and the evaluation of alternatives.

This community of practice does not replace other educational approaches but complements and adds dynamics never before considered.

One critical factor for the success of this proposal is an adequate teacher training in ICT and its modalities and to provide students with self-training contexts that promote collaborative practices between teachers and students.

The objectives that are considered essential to a community of practice are:

1. To promote cooperation and collaboration among participants on educational topics;
2. To promote the exchange of ideas to support school practices in classroom settings;
3. To encourage communication across the different thematic forums promoting educational issues and reflection;

4. To encourage participation in the Facebook® forum as a means of social interaction developing friendship and cooperation among participants in the community.

7 Conclusion

We think that the values that strengthen the community sense like friendship, respect, participation, solidarity, common good, and the feeling of belonging to a community must be encouraged. The peers' community is the first kind of community encountered by the teenager when he/she tries to be independent from his/her family, where he/she shares values, objectives and interests. When individuals come together to form a group, they resonate with each other, producing through their communication a social attractor and a social hierarchy characterized by a leadership structure, the rules to be followed, and the sanctions to be assumed provided that the members are accepted in the group. In fact, under the surface of the social, individual resonances harmonize in the sense that people develop a shared view. Moreover, in the same school different opposing subgroups coexist that try to differentiate because of its strength, age, experience, abilities and daring. Therefore, a strategic policy for educational intervention should aim to transform the rivalry among these subgroups into a cooperative force with common interests at a higher level through computer-aided communication. This would enhance the values in the peers' subculture toward the development of social attractors that would change the social pattern toward a cooperative one at a school level [24]. This transformation could be implemented through computer-aided intervention strategies designed to make use of communication-oriented Internet social networks sites like Facebook®.

Acknowledgement

This work was supported in part by the project PAPIIT IN105909, Universidad Nacional Autónoma de México and project CONACYT 152008.

References

[1] Velázquez-Guzmán, M.G. and Lara-Rosano, F. (2009). Social Development for Living Together: School Practice Diagnosis and Systemic Proposals. In Personal and Spiritual Development in the World of Cultural Diversity, Vol. VI, G.E. Lasker and K Hiwaki (eds), IIAS, Windsor, Canada, pp 31-36.

[2] Stacey, R.D. (2001). Complex Responsive Processes in Organizations, Routledge, London.

[3] Marion, M. (1995). The Edge of Organization: Chaos and Complexity Theories of Formal Social Systems, Sage Publications Inc., Thousand Oaks, CA

[4] Mead, G.H. (1934). Mind, Self and Society. University of Chicago Press, Chicago.

[5] Muuss, R (1980). Adolescent Behavior and Society. Random Hansom, New York.

[6] Handel, G. et al. (2008) Children and Society. Roxbury Publishing Co., Los Angeles, CA.

[7] Savin, R. (1980). An Ethological Study of Dominance Formation and Maintenance in a Group of Human Adolescents. In Muuss, R Adolescent Behavior and Society. Random Hansom, New York.

[8] Corsaro, W. (1985) Friendship and Peer Culture in the Early Years. Ablex Publishing Corporation, Norwood, NJ.

[9] Velázquez-Guzmán, M.G. and Lara-Rosano, F. (2007). The Complexity of Building up Communitarian Coexistence in the School: A Hermeneutical Approach, ActaSystemica, Vol 7, No 2, pp 17-24.

[10] McGregor, D. (1960). The Human Side of Management. McGraw-Hill, New York.

[11] Velázquez-Guzmán, M.G. and Escobedo, C.I. (2008). Agresores, Agredidos y Mediadores. Universidad Pedagógica Nacional, México.

[12] Fromm, E. (2006) Anatomía de la destructividadhumana. México: Siglo XXI .

[13] Sherif, M. and Sherif, C. (1964). Problems of Youth. Aldine Pub. Co., Chicago.

[14] Baxter, L.A. and Montgomery, B.M. (1996). .Relating: Dialogues and Dialectics. Guilford Press, New York.

[15] Gadamer, H. G. (1990). Wahrheit und Methode I. J.C.B. Mohr (Paul Siebeck), Tübingen.

[16] Wenger, E. (1998). Communities of practice - Learning, Meaning and Identity. Cambridge: Cambridge University Press.

[17] Wenger, E. (2000). Communities of practice and social learning systems. Organization, Vol. 7, No 2, pp 225-256.

[18] [18] Wenger, E. (2001). Supporting communities of practice. A survey of community oriented technologies, San Juan: North.

[19] Miranda-Pinto, M.S. and Osorio A.J. (2010). Colaboración y aprendizaje en el ciberespacio. Revista Mexicana de Investigación Educativa, Vol. XV, No. 44, pp. 35-64.

[20] Subrahmanyam, K. and Greenfield, P. (2008). Online Communication and Adolescent Relationships. The Future of Children, Vol.18, No. 1, Spring, pp 119 – 146.

[21] Prigogine, I. (1997). The End of Certainty: Time, Chaos and the New Laws of Nature. The Free Press, New York.

[22] Langton, C. (1989). Artificial Life. In Nadel, L. and Stein, D. (eds). Lectures in Complex Systems. Addison-Wesley, Reading, Mass.

[23] National School Boards Association (2007). Creating and Connecting: Research and Guidelines on Online Social –and Educational- Networking, NSBA org, July.

[24] Velázquez-Guzmán, M.G. and Lara-Rosano, F. (2010). Teenagers Educational Intervention through the Peers' Subculture. In Personal and Spiritual Development in the World of Cultural Diversit,y Vol VII, G.E. Lasker and K. Hiwaki (eds), IIAS,, Windsor, Canada, pp 27-32.

A Market Based Approach for Sensor Resource Allocation in the Grid

Li Chunlin, Huang Jin Hui and Li LaYuan
Department of Computer Science, Wuhan University of Technology, Wuhan 430063, P.R.China
E-mail: chunlin74@yahoo.com.cn, jwtu@public.wh.hb.cn

Keywords: sensor, grid computing, market

Received: August 5, 2011

The paper studies a market based approach for sensor resources allocation in sensor enabled grid computing environment. The paper presents an efficient mechanism to assign sensor resources to appropriate sensor grid users on the basis of negotiation results among participants. We model the sensor allocation problem by introducing the sensor utility function. The goal is to find a sensor resource allocation that maximizes the total profit. The paper proposes a distributed optimal sensor resource allocation algorithm. The performance evaluation of proposed algorithm is evaluated and compared with other resource allocation algorithm for sensor grid.

Povzetek: Predstavljena ja postavitev senzorjev v omrežje na osnovi pogajanj med udeleženci.

1 Introduction

Grid computing is based on the concept of the coordinated sharing of distributed and heterogeneous resources to solve large-scale problems in dynamic virtual organizations. The grid computing paradigm can be extended to include the sharing of sensor resources in sensor networks. Integrating sensor networks and grid computing in sensor-grid computing is like giving ‘eyes’ and ‘ears’ to the computational grid. Real-time information about phenomena in the physical world can be processed, modeled, correlated and mined to permit on-the-fly decisions and actions to be taken on a large scale [1, 16]. Sensor grids extend the grid computing paradigm to the sharing of sensor resources in wireless sensor networks. By combining the complementary strengths of sensor networks and grid computing, sensor grids can support applications that require real-time information from the physical environment and vast amount of computational and storage resources. Examples include environment monitoring with prediction and early warning of natural disasters, and missile detection, tracking and interception [3].

However, sensor enabled grid is a widespread distributed system and maybe consists of many sensors belonging to individual user who does not know with other users at all. They would like to do the things that benefit themselves most, which means they are rational and selfish. Therefore, sensor grid needs to provide incentives to encourage every user to contribute their resources to other users. Sensor grids being a relatively new area of research, there are many issues left unaddressed regarding their design. One of the major challenges in the design of sensor enabled grid is how to efficiently schedule sensor resource to user jobs across the collection of sensor resources. Due to the energy limitation and also to prolong the lifetime of the sensor grid, conservation of energy consumed is an important consideration in managing sensor grids. Making decisions on how best to utilize limited sensor resources in order to satisfy grid users’ demands without conflict and without wasting resources is the key issue in sensor grid. The resource allocation in grid computing systems has been extensively studied in the past. There are some important differences between sensor resource and computational resource. Thus, existing allocation algorithms for traditional grid environment may not work well in sensor enabled grids.

In this paper, we present a market based approach for sensor resources allocation in sensor enabled grid computing environment. Since sensor users’ tasks might compete for the exclusive usage of the same sensing resource we need to allocate individual sensors to sensor users’ tasks. Sensor grid tasks are usually characterized by an uncertain demand for sensing resource capabilities. We model this allocation problem by introducing the sensor utility function. The goal is to find a sensor resource allocation that maximizes the total profit. The paper proposes a distributed optimal sensor resource allocation algorithm. The performance evaluation of proposed algorithm is evaluated and compared with other resource allocation algorithm in sensor grid. The paper also gives the application example of proposed approach.

The rest of the paper is structured as followings. Section 2 discusses the related works. Section 3 presents a market based approach for sensor resource allocation in the grid. Section 4 describes sensor resources allocation algorithm. In section 5 the experiments are conducted and discussed. Section 6 gives the conclusions to the paper.

2 Related Works

There are certain researches aiming to combining grid environments with wireless sensor network [1~17], which incorporate sensors into the existing grid systems as the consumers of grid resources and provide sensor services to other grid nodes. Peter Komisarczuk et al. [1] discusses research direction in an Internet sensor grid for malicious behavior detection, analysis and countermeasures. They outline some experiences with these sensors and analyzing network telescope data through Grid computing as part of an “intelligence layer” within the Internet. M. Pallikonda Rajasekaran et al. [2] propose a wireless sensor grid architecture for monitoring the health status of different groups of patients to provide a platform for physicians and researchers to share information with distributed database and computational resources to facilitate analysis, diagnosis, prognosis and drug delivery. Hock Beng Lim et al. [3] design an integrated and flexible scheduler for a sensor grid testbed based on the SPRING framework. Several scheduling and load balancing algorithms were implemented within this scheduler to suit the unique characteristics of sensor jobs. The scheduler can use an appropriate scheduling or load balancing algorithm to suit the requirements of the resource owner and users. Nikolaos Preve et al. [4] proposed the sensor grid enhancement data management system, called SEGEDMA ensuring the integration of different network technologies and the continuous data access to system users. The main contribution is to achieve the interoperability of both technologies through a novel network architecture. Huang-Chen Lee et al. [5] discuss the considerations of designing a low-cost WSN-based rain gauge grid, which provides high resolution mapping of precipitation. Preliminary experimental results are presented. Fox, G. et al. [6] propose a collaborative sensor grid framework to support the integration of a sensor grid with collaboration and other grids. The framework includes a grid builder tool for discovering and managing grid services and remote, distributed sensors. It provides a real-time collaborative client to enable distributed stakeholders to have a consistent view of displayed sensor streams. They illustrate the versatility of the framework by constructing a robot based customizable application for shared situational awareness. Hock Beng Lim et al. [7] aim to build a large-scale sensor grid infrastructure that can seamlessly integrate heterogeneous sensor resources from different projects distributed across a wide geographical area. Sanabria, J. et al. [8] discussed a deployment framework, which leverages on existing grid computing technologies to provide middleware that integrates wireless sensor networks and grid infrastructures. They demonstrated the work on enabling a sensor grid infrastructure for acoustic surveillance applications. Hock Beng Lim et al. [9] developed a sensor grid architecture framework, called the Scalable Proxy-based aRchItecture for seNsor Grid (SPRING). They designed the National Weather Sensor Grid (NWSG), a large-scale cyber-sensor infrastructure for environmental monitoring. The NWSG integrates mini weather stations deployed geographically across Singapore for weather data collection, processing and management. Toshihiro Matsui et al. [10] deal with distributed resource allocation problem for distributed sensor networks. Distributed optimization algorithm is used to understand the problems and to design the cooperative protocols. A distributed cooperative observation system using agency model has been developed. Yan YuJie et al. [11] described the architecture of wireless sensor grid, also designed a connecting platform named MPAS. The advantage of MPAS is that it is based on Web service resource framework, with the ability of integrating multiple sensor networks with grid; also it can actuate sensor network and support interoperability among multiple sensor networks. Mohammad Mehedi Hassan et al. [12] discuss one of the most important issues in Sensor-Grid, i.e., to develop a fast and flexible content-based publish/subscribe information dissemination (CBPSID) system for automatic fusion, interpretation, sharing and delivery of huge sensor data to consumers as the entire Sensor-Grid environment is very dynamic.

Se-Jin Oh et al. [13] present the design and implementation a u-Healthcare SensorGrid gateway to connect transparently a sensor network and a Grid network for providing convenient and speedy u-Healthcare services to users. They implement a mobile monitoring system for monitoring patient’s status by using a mobile device such as PDA. Xiaolin Li et al. [14] propose an autonomic management framework (ASGrid) to address the requirements of emerging large-scale applications in hybrid grid and sensor network systems. They proposed the autonomic sensor grid system concept in a holistic manner targeted at non-trivial large applications. Toshihiro Matsui et al. [15] propose a model of distributed resource allocation problem for distributed sensor networks. Several models based on constraint network and another model based on concept of agency, are compared. The constraint network formalization which are similar to resource allocation problem of agency model, is shown. Yong-Kang Ji et al. [17] discuss the selfish problems of Sensor Web and resolve them using specific designed mechanism, and describe several scenarios of the applications in Sensor Web. The works [18~22] mainly deal with resource allocation, QoS optimization in the computational grid and don’t consider exploiting services of sensor to support sensor enabled grid. The differences between paper [20] [21] and this paper are that this paper deals with optimal sensor resources allocation, paper [20] considers two level market solution for optimal resource scheduling in computational grid; paper [21] considers multiple QoS optimization in computational grid. Computational grid is different from sensor grid. In sensor grid environment, sensor nodes have present severe limitations in terms of processing, memory capabilities and energy, but computational grid doesn’t have energy problem, and also have enough processing, memory capabilities. Considering all these limitations of sensor nodes, it is important for the sensor grid system to manage energy consumption without compromising system's performance. Our proposed optimization in sensor grid considers energy consumption in the sensor node. The objective of the paper is to maximize the utility of the system without exceeding the total energy available, the expense budget, and deadline.

3 A Market based Approach for Sensor Resource Allocation in the Grid

3.1 System Model

The paper formulates optimal sensor resources allocation in sensor enabled grid computing environment by adopting computational economy framework. The proposed model consists of two types of agents: the sensor resource agents that represent the economic interests of the underlying sensor resources providers of the sensor grid, the sensor user agents that represent the interests of grid user using the grid to achieve goals. Interactions between the two agent types are mediated by means of market mechanisms. Market mechanism in economics is based on distributed self-determination, the variation of price reflects the supply and demand of resources, and market theory in economics provides precise depiction for efficiency of resource scheduling. Sensor user agents are allowed to specify their requirements and preference parameters by a utility model. As a result, a market based sensor grid model inherently supports sensor users with diverse requirements for the execution of their sensor jobs. The utility values are calculated by the supplied utility function that can be formulated with the sensor job parameters. The request is analyzed by the scheduler of grid market. Whenever a new sensor user agent is created, it is first given an endowment of electronic cash to spend to complete its sensor job. A sensor job can be characterized by deadline, budget, and sensor task requirements. The sensor grid market mechanism allows multiple sensor resource agents and sensor user agents to negotiate simultaneously, it uses price-directed approach to allocate appropriate sensor resources. In this price-directed approach, an initial set of prices is announced to the sensor user agent. Sensor users could update their allocations based on the sensor provider's price policy, and iteratively approach an optimal solution. In each iteration, sensor user agents individually determine their optimal allocation and communicate their results to the sensor resource agents. Sensor resource agents then update their prices and communicate the new prices to the sensor user agents and the cycle repeats. Prices are then iteratively changed to accommodate the demands for resources until the total demand equals to the total amount of sensor resources available. Sensor resource agents publish sensor descriptions to the market. Sensor providers compete actively for sensor jobs from sensor users and execute them for gaining profits. Every sensor provider tries to maximize its profit based on its resource capability and energy consumption. We assume that the sensor resource agents do not cooperate. Instead, they act non-cooperatively with the objective of maximizing their individual profits. The sensor resource agents compete among each other to serve the sensor user agents. The sensor user agents do not collaborate either, and try to purchase as much sensor resource as possible with the objective of maximizing their net benefit.

3.2 Mathematic Formulation

In this section, we set up the mathematical models for optimal sensor resources allocation in sensor enabled grid computing environment.

First gives notations to be used in the following sections:

p_j : the price of the sensor resource unit in sensor *j*.

B_i : the expense budget given to a sensor grid application *i*

E_j : the energy limit of sensor resource *j*

u_i^j : money paid to the sensor resource *j* by sensor grid user *i*

q_i^n : sensing task of *i*th sensor grid application's *n*th sensor job

t_i^n : the time taken by sensor grid application *i* to complete sensor job *n*

T_i : time limits given by sensor grid application *i* to complete all sensor jobs

es_j : energy consumption of unit resource of sensor *j*

ec_i^j : energy dissipation of sensor resource *j* for completing sensor grid application *i*

x_i^j : sensor allocated to sensor grid application *i* by the sensor resource *j*

It is assumed that the sensor grid system consists of multiple grid sites that contain sensor nodes and ordinary fixed grid nodes. Sensor nodes consist of a collection of sensors *M* connected by sensor network. In sensor grid, an application set is denoted as $App = \{App_1, App_2 ... App_i\}$ and a sensor resource set is denoted as $s = \{s_1, s_2, s_j ...\}$, c_j is the available capacity of the sensor resource s_j . Sensor node estimates its energy consumption rate es_j for executing the application set, and the energy constraint is E_j . If the energy consumption is proportional to sensor resource unit, as is the case with battery energy. x_i^j is sensor allocated to sensor grid application *i* by the sensor resource *j*. The energy limitation of the sensor imposes a constraint as follows:

$$\sum_{i=1}^{I} ec_i^j = \sum_{i=1}^{I} es_j x_i^j \le E_j \qquad (3.1)$$

The sensor jobs are assumed to acquire real time data from sensor grid node, mutually independent, and can be executed at certain sensor grid node. As soon as a sensor job arrives, it must be assigned to one sensor grid node for processing. When a sensor job is completed, the executing sensor node will return the results to the originating sensor nodes or ordinary fixed grid node of the job. We use *SJ* to denote the set of all sensor jobs generated by sensor grid application *i*, $SJ_i = \{SJ_i^1, SJ_i^2 ... SJ_i^n\}$. Each sensor job can be described as $SJ_i^n = (t_i^n, q_i^n)$, in which t_i^n stands for the time taken by the *i*-th sensor grid application to complete *n*-th sensor job, q_i^n stands for sensing task of *i*th sensor user' s *n*th sensor job. There are no dependencies among the sensor jobs, so the submission order and completion order won't impact on the execution result.

We assume that sensor user has an associated utility function $U_i(x_i^j)$ if the allocated sensor resource is x_i^j . Now, we formulate the problem of optimal sensor resources allocation in sensor enabled grid environment as constraint optimization problem, the utility of the sensor grid system is defined as the sum of sensor grid user utilities.

x_i^j is sensor resource allocation obtained by sensor user *i* from the sensor provider *j*. The utility function for sensor user application A_i depends on allocated sensor resources x_i^j . The objective of sensor resource allocation optimization is to maximize the utility of the sensor grid system without exceeding the sensor capacity, the energy limit of sensor, expense budget and the deadline. We formalize the problem using nonlinear optimization theory; the sensor resource allocation optimization in sensor grid can be formulated as follows.

$$Max\, U_{SensorGrid} = \sum_{i=1}^{I} U_i(x_i^j) \tag{3.2}$$

Subject to $B_i \geq \sum_j u_i^j$, $T_i \geq \sum_{n=1}^{N} t_i^n$,

$$\sum_{i=1}^{I} ec_i^j \leq E_j \, , c_j \geq \sum_i x_i^j \tag{3.3}$$

Equation 3.2 is the sensor grid system utility maximization formulation. The sensor grid utility is defined as the sum of utilities for all sensor users. The cost overhead accrued to complete jobs cannot exceed the expense budget B_i . The time for completing all sensor jobs of user application *i* cannot exceed the deadline T_i . The total energy consumed by sensor *j* for executing sensor user applications cannot exceed an energy limits E_j . Aggregate allocated resource does not exceed the total sensor capacity c_j .

We can apply the Lagrangian method to solve such a problem. The Lagrangian approach is used to solve constrained optimization problems. Let us consider the Lagrangian form of sensor resource allocation optimization problem in sensor grid:

$$L(\lambda_i, \beta_i, \varphi_i, \gamma_i) = \sum_i U_i(x_i^j) - \lambda_i(\sum_j u_i^j - B_i) - \beta_i(\sum_n t_i^n - T_i) - \varphi_i(\sum_i ec_i^j - E_j) - \gamma_i(\sum_i x_i^j - c_j) \tag{3.4}$$

$$\partial L / \partial x_i^j = \partial U_i(x_i^j) / \partial x_i^j - \gamma_i - \varphi_i \, \partial ec_i^j / \partial x_i^j - \beta_i \, \partial t_i^n / \partial x_i^j$$

$$x_i^j (\partial U_i(x_i^j) / \partial x_i^j - \gamma_i - \varphi_i \, \partial ec_i^j / \partial x_i^j - \beta_i \, \partial t_i^n / \partial x_i^j) = 0 \tag{3.5}$$

$$\gamma_i (\sum_i x_i^j - c_j) = 0,$$

$$\beta_i (\sum_n t_i^n - T_i) = 0$$

$$\varphi_i (\sum_i ec_i^j - E_j) = 0 \tag{3.6}$$

Where λ_i , β_i , γ_i is the Lagrangian multiplier of sensor user *i*. Thus, given that the sensor grid knows the utility functions $U_i(x_i^j)$ of all sensor user *i*, this optimization problem can be mathematically tractable. However, in practice, it is not likely to know all the $U_i(x_i^j)$, and it is also infeasible for sensor grid environment to compute and allocate sensor resource in a centralized fashion. Solving the objective function $Max\, U_{SensorGrid}$ requires global coordination of all sensor users, which is impractical in distributed environment such as the sensor grid.

The system model presented by (3.2) is a nonlinear optimization problem with *N* decision variables. Since the Lagrangian is separable, the maximization of the Lagrangian can be processed in parallel for sensor users and sensor providers respectively. The sensor resource allocation $\{x_i^j\}$ solves problem (3.2) if and only if there exist a set of nonnegative shadow costs $\{\gamma_i\}$. Generally solving such a problem by typical algorithms such as steepest decent method and gradient projection method is of high computational complexity, which is very time costing and impractical for implementation. In order to reduce the computational complexity, we decompose the sensor utility optimization problem (3.2) into two subproblems for sensor users and sensor providers. The shadow costs suggest a mechanism to distribute the sensor resource optimization between the sensor users and the sensor grid. We consider the Lagrangian multipliers γ_i to be the prices charged by sensor resource providers in sensor market, equation (3.7) describes sensor user's behavior in sensor market as a sensor consumer, and equations (3.8) describe the sensor provider's strategy as a sensor supplier. By decomposing the Kuhn-Tucker conditions into separate roles of sensor user and sensor supplier at sensor market, the centralized problem (3.2) can be transformed into a distributed problem. Sensor user's payment is collected by the sensor providers. The payments of sensor users paid to sensor providers are the payments to resolve the optimality of sensor resource allocation in the sensor market. We decompose the problem into the following two subproblems (3.7) which is sensor user optimization problem and (3.8) which is sensor providers optimization problem, seek a distributed solution where the sensor provider does not need to know the utility functions of individual sensor user. Two maximization subproblems correspond to sensor user optimization problem as denoted by (3.7)

$$SA = Max\{(B_i - \sum_j u_i^j) + (T_i - \sum_{n=1}^{N} t_i^n)\} \quad s.t. T_i \geq \sum_{n=1}^{N} t_i^n \tag{3.7}$$

and a sensor provider optimization problem as denoted by (3.8) .

$$SR = Max \sum u_i^j \log x_i^j + (E_j - \sum_{i=1}^{I} ec_i^j) \quad s.t.\ c_j \geq \sum_i x_i^j, \sum_{i=1}^{I} ec_i^j \leq E_j \tag{3.8}$$

In (3.7), for the sensor user optimization problem, the sensor user gives the unique optimal payment to sensor provider under the deadline constraint to maximize the sensor user's benefit. $(B_i - \sum_j u_i^j)$ represents the money surplus of sensor user, which is obtained by budgets subtracting the payments to sensor providers. So, the objective of (3.7) is to get more surpluses of money at the some time complete jobs for sensor user as soon as possible. In (3.8), for the sensor provider's optimization problem, different sensor providers compute optimal sensor resource allocation for maximizing the revenue of their own and minimizing the energy consumption for completing sensor grid application. We could have chosen any other form for the utility that increases with x_i^j . But we chose the *log* function because the benefit increases quickly from zero as the total allocated sensor resource increases from zero and then increases slowly. Moreover, *log* function is analytically convenient, increasing, strictly concave and continuously differentiable. The benefits of sensor provider are affected by payments of sensor users, allocated resources and energy consumption. It means that the revenue increases with allocated sensor resources increasing and payment increasing, also with energy consumption decreasing. The objective of sensor providers is to maximize $u_i^j \log x_i^j$ and minimize $\sum_{i=1}^{I} ec_i^j$ under the constraints of their energy limit. Sensor providers can't dissipate energy more than E_j, which is the upper limit of energy presented by sensor providers. $(E_j - \sum_i ec_i^j)$ represents the energy surplus of sensor provider which is obtained by the energy limit subtracting energy dissipation. Thus, the optimization framework provides a distributed approach to the sum utility maximization problem. Sensor user layer problem adaptively adjusts sensor user's resource demand based on the current sensor resource conditions, while the sensor resource layer adaptively allocates sensor required by the upper layer. The interaction between the layers is now controlled through the use of the variable γ_i, which is the price charged from sensor users by sensor provider and coordinates the sensor user demand and the supply of sensor resource.

For the sensor user optimization problem, the sensor user gives the unique optimal payment to sensor provider under deadline constraint to maximize the sensor user's satisfaction.

In equation (3.7), Let p_j denote the unit price of sensor *j*, Let the pricing policy, $p = (p_1, p_2, \cdots, p_j)$, denote the set of sensor prices of all the sensor providers at the sensor resource layer. The sensor user *i* receives sensor resource proportional to its payment relative to the sum of the sensor provider's revenue. Let x_i^j be the sensor resource units allocated to sensor user *i* by sensor provider *j*.

The time taken by the *i*th sensor user to complete *n*th sensor job is:

$$t_i^n = \frac{q_i^n p_j}{c_j u_i^j} \tag{3.9}$$

We reformulate sensor user optimization problem as

$$Max\{(B_i - \sum_j u_i^j) + (T_i - \sum_{n=1}^{N} \frac{q_i^n p_j}{c_j u_i^j})\} \tag{3.10}$$

We take derivative and second derivative of *SA* with respect to u_i^j :

$$SA'(u_i^j) = \frac{dSA}{d\,u_i^j} = \sum_{n=1}^{N} \frac{q_i^n\ p_j}{(u_i^j)^2 c_j} - 1$$

$$SA''(u_i^j) = \frac{d^2 SA(u_i^j)}{d\,(u_i^j)^2} = -\sum_{n=1}^{N} \frac{q_i^n p_j}{(u_i^j)^3 c_j}$$

$SA''(u_i^j) < 0$ is negative due to $u_i^j > 0$ .The extreme point is the unique value maximizing the sensor user's utility under completed time limits. The Lagrangian for the sensor user's utility is $L(u_i^j)$.

$$L(u_i^j) = (B_i - \sum_j u_i^j) + (T_i - \sum_{n=1}^{N} \frac{q_i^n p_j}{c_j u_i^j}) + \lambda(T_i - \sum_{n=1}^{N} t_i^n) \tag{3.11}$$

Where λ is the Lagrangian constant. From Karush-Kuhn-Tucker Theorem we know that the optimal solution is given $\partial L(u_i^j) / \partial u_i^j = 0$ for λ >0.

$$\partial L(u_i^j) / \partial u_i^j = -1 + \frac{q_i^n\ p_j}{c_j (u_i^j)^2} + \lambda \frac{q_i^n\ p_j}{c_j (u_i^j)^2} \tag{3.12}$$

Let $\partial L(u_i^j) / \partial u_i^j = 0$ to obtain

$$u_i^j = \left(\frac{(1+\lambda)q_i^n\, p_j}{c_j}\right)^{1/2} \quad (3.13)$$

Using this result in the constraint equation $T_i = \sum_{n=1}^{N} t_i^n$, we can determine $\theta = 1+\lambda$ as

$$T_i = \sum_{n=1}^{N} t_i^n = \sum_{n=1}^{N} \frac{q_i^n p_j}{c_j u_i^j} = \sum \frac{q_i^n p_j}{c_j} \left(\frac{\theta q_i^n\, p_j}{c_j}\right)^{-1/2} \quad (3.14)$$

We obtain

$$(\theta)^{-1/2} = \frac{T_i}{\sum_{k=1}^{N} \left(\frac{p_k\, q_i^k}{c_j}\right)^{1/2}} \quad (3.15)$$

We get u_i^{j*}

$$u_i^{j*} = \left(\frac{q_i^n\, p_i}{c_j}\right)^{1/2} \frac{\sum_{k=1}^{N} \left(\frac{q_i^k\, p_k}{c_j}\right)^{1/2}}{T_i} \quad (3.16)$$

u_i^{j*} is the unique optimal solution to sensor user optimization problem. It is the optimal payment of sensor user *i* to sensor provider *j* under the completion time constraint to maximize the sensor user's benefits.

For the sensor provider's optimization problem, different sensor providers compute optimal sensor allocation for maximizing the revenue of their own under constrains of energy limit.

In equation (3.8), $\sum u_i^j log\, x_i^j$ presents the revenue obtained by sensor resource *j* from sensor users. The energy consumption of sensor provider for executing users' task can't exceed more than E_j , which is the upper limit of sensor power.

The energy dissipation of sensor resource *j* for completing grid application *i* be written:

$$ec_i^j = es_j x_i^j \quad (3.17)$$

We reformulate sensor provider's optimization problem as

$$SR = Max \sum u_i^j \log x_i^j + (E_j - \sum_{i=1}^{I} es_j x_i^j) \quad (3.18)$$

We take derivative and second derivative with respect to x_i^j :

$$SR'(x_i^j) = \frac{u_i^j}{x_i^j} - es_j \ , \ \ SR''(x_i^j) = -\frac{u_i^j}{x_i^{j\,2}}$$

$SR''(x_i^j) < 0$ is negative due to $0 < x_i^j$ .The extreme point is the unique value maximizing the revenue of sensor provider. The Lagrangian for *SR* is $L(x_i^j)$

$$L(x_i^j) = \sum_{i=1}^{I} u_i^j \log x_i^j + (E_j - \sum_{i=1}^{I} es_j x_i^j) + \delta(E_j - \sum_{i=1}^{I} es_j x_i^j)$$
$$= \sum_{i=1}^{I} (u_i^j \log x_i^j - (\delta+1) es_j x_i^j) + (\delta+1) E_j \quad (3.19)$$

Where σ is the Lagrangian constant. From Karush-Kuhn-Tucker Theorem we know that the optimal solution is given $\partial L(x) / \partial x = 0$ for λ >0.

$$\partial L(x_i^j) / \partial x_i^j = \frac{u_i^j}{x_i^j} - (\delta + 1) es_j \quad (3.20)$$

Let $\partial L(x) / \partial x = 0$ $\qquad x_i^j = \frac{u_i^j}{(\delta+1) es_j}$

(3.21)

Using this result in the constraint equation, we can determine $\sigma = 1+\delta$ as

$$E_j = \frac{1}{\sigma} \sum_{i=1}^{n} u_i^j \quad , \quad \sigma = \frac{\sum_{k=1}^{n} u_k^j}{E_j}$$

We obtain

$$x_i^{j*} = \frac{u_i^j E_j}{es_j \sum_{k=1}^{n} u_k^j} \quad (3.22)$$

x_i^{j*} is the unique optimal solution to sensor provider's optimization problem. It means that sensor providers allocate x_i^{j*} to sensor user to maximize its revenue.

4 Description of Algorithms

The objective of market based sensor allocation is to maximize the utility of the sensor grid system without exceeding the energy limit, expense budget and the deadline. The proposed algorithm decomposes optimal sensor resource allocation optimization problem into a sequence of two sub-problems via an iterative algorithm. In each iteration, in sub problem 1, the sensor user computes the unique optimal payment to sensor provider under expense budget and the deadline constraint to maximize the sensor user's benefit. The sensor user individually solves its fees to pay for sensor resource to complete its all sensor jobs, adjusts its sensor demand and notifies the sensor provider about this change. In sub problem 2, different sensor providers compute optimal sensor resource allocation for maximizing the revenue of their own under energy constraint. Sensor provider updates its price according to optimal payments from sensor user, and then sends the new prices to the sensor user s and allocates the sensor resource for sensor user, and the cycle repeats. The iterative algorithm that achieves sensor resource allocation optimization is described as follows.

Algorithm 1 Market based Sensor Allocation Algorithm (MSA)

Sensor user *i*

Receives the new price p_j from the sensor resource *j*;

$u_i^{j*} = Max\, U_{SA}(u_i^j)$; //calculates u_i^{j*} to maximize $U_{SA}(u_i^j)$

If $B_i \geq \sum_j u_i^j$

Then Return u_i^{j*} to sensor resource *j* ;

Else Return Null;

Sensor resource *j*

Receives optimal payments u_i^{j*} from sensor user *i*;

$x_i^{j(n+1)*} = Max\, U_{SR}(x_i^j)$; // Calculates its optimal sensor resource $x_i^{j(n+1)*}$

$p_j^{(n+1)} = \max\{\varepsilon, p_j^{(n)} + \eta(x^j p_j^{(n)} - c_j)\}$; // Computes a new price

// $x^j = \sum_i x_i^j$, $\eta > 0$ is a small step size parameter, *n* is iteration number

Return sensor resource price $p_j^{(n+1)}$ to all sensor users;

5 Simulations

In this section, the performance evaluation of our mechanism is conducted. The sensor grid simulator is implemented on top of the JAVASIM network simulator [22]. In the experiments, 150 sensors are uniformly deployed in a field that is $500m \times 500m$ in area. There are also 16 base stations that are deployed based on a uniformly random distribution. Sensor tasks are created in uniformly distributed locations in the field. There are a total of 150 sensor resources and 600 sensor users are taken for experimental evaluation of the system. Energy consumption is represented as a percentage of the total energy required to execute all job and meet deadlines. Assume that the maximum power, $P_{\max}$, corresponds to running all jobs with the maximum processing frequency. The maximum frequency is assumed to be $f_{\max} = 1$ and the maximum frequency-dependent power is $P_{\max} = 1$. When the power capacity for each interval is limited, we can only consume a fraction of $P_{\max}$ when processing requests during a given interval. Jobs arrive at each site s_i, *i*=1, 2,…, *n* according to a Poisson process with rate α. The energy cost can be expressed in grid dollar that can be defined as unit energy processing cost. The initial price of energy is set from 10 to 500 grid dollars. Sensor users submit their jobs with varying deadlines. The deadlines of sensor user application are chosen from 100ms to 400ms. The budgets of sensor applications are set from 100 to 1500 grid dollars. Each experiment is repeated 6 times and 95% confidence intervals are obtained.

The experiments are conducted to compare proposed market based sensor allocation algorithm (MSA) with an integrated and flexible scheduler for a sensor grid [3], which makes use of several scheduling and load balancing algorithms to suit the characteristics of sensor jobs in sensor grid environment. The reason for choosing reference [3] as the comparison is that both our work and reference [3] provide a resource scheduling and allocation algorithm for sensor grid environment. In [3], they adapted four existing multiprocessor scheduling algorithms to suit the sensor grid scenario: Earliest Deadline First (EDF), First Come First Served (FCFS), Least Laxity First (LLF), and Shortest Job Next (SJN). EDF is a scheduling algorithm that allocates higher priorities to jobs closer to their deadline. As an alternative to the static priority schedulers, it guarantees schedulability when node utilization is full. In the case where deadlines are equal, we modified the algorithm to use the duration of the job as a tie breaker, with shorter duration jobs given a higher priority. FCFS is the simplest of the four scheduling algorithms. Using a queue structure, this algorithm simply adds new jobs to the end of the queue as they arrive, picking jobs for execution only from the front of the queue. A derivative of EDF, LLF is a scheduling algorithm that allocates priorities on the amount of laxity a job has: the lower the laxity, the higher the priority. The laxity, or slack time, is the time left until its deadline after the job is completed, assuming that the job could be executed immediately. SJN allocates priorities statically depending on the duration of jobs. Shorter jobs are favored, and are given higher priorities. Similar to LLF, the modification made to the algorithm was to allocate a higher priority to the job with the nearer deadline when durations are equal.

The paper performed the evaluations to measure the impact of the algorithms using several performance metrics, and test whether the algorithm can satisfy performance objectives such as sensor resource utilization, allocation efficiency, sensor job execution success, response time and energy consumption ratio. We study the performances of proposed market based sensor allocation algorithm (MSA) with several scheduling and load balancing algorithms for a sensor grid [3]. We choose the performance metrics and simulation parameter according to the reference [3]. Performance metrics include in terms of allocation efficiency, execution success ratio, response time and energy consumption ratio, resource utilization ratio. Allocation efficiency is defined as the percentage of allocated sensor among total available sensor resources. Execution success ratio is the percentage of sensor jobs executed successfully before their deadline. Energy consumption ratio is defined as the percentage of consumed energy of sensors to complete the jobs. We compare the algorithms by varying load factor to study how they affect the performance of the algorithms.

The following four figures (Figs.1, Figs.2, Figs.3, and Figs.4) are to study resource allocation efficiency, execution success ratio, response time, energy consumption ratio and resource utilization ratio under different load factor (LF) respectively. Load factor varies from 0.1 to 0.9. Fig.1 shows the effect of load factor on allocation efficiency. When LF=0.5, allocation efficiency of MSA is as much as 17% less than that with LF=0.1. The allocation efficiency is larger when the load factor (LF) is smaller. When load factor increases, system load increases; some sensor user agent's requirements can't be processed on time. The sensor user agents with low budget don't have enough money to buy sensor and can't complete jobs before deadline; this leads to low allocation efficiency. When load factor is 0.5, allocation efficiency of MSA is 27% more than FCFS. Compared with FCFS and SJN, the allocation efficiency of MSA decreases more slowly when the load factor increases. When the load factor is 0.6, the allocation efficiency of LLF decreases to 54%, the allocation efficiency of MSA decreases to 84%. The allocation efficiency of MSA is better than SJN, LLF and FCFS. Considering the execution success ratio, from the results in Fig.2, when load factor is 0.5, the execution success ratio of SJN is 28% less than that using MSA. When load factor increases, execution success ratio of SJN and FCFS deteriorate quickly. SJN and FCFS scheduling algorithms don't consider optimization of both sensor resource providers and sensor users, it wants to minimize the runtime of sensor jobs. EDF has higher execution success ratio than MSA. When the load factor increases, fewer requests from sensor user agents can be admitted into the system due to the increase of system burden, so, fewer requests from sensor user agents can be executed successfully before their deadline. Fig.3 shows the energy consumption ratio under different load factors. When load factor increases, more requests need to be processed within one interval and the energy consumption ratio increases. When increasing the load factor by LF=0.7, the energy consumption ratio of MSA is as much as 25% more than LF=0.4. Under same load factor (LF=0.8), the energy consumption ratio of MSA is 16% less than that of EDF. The energy consumption ratio of MSA and EDF is less than that of SJN, LLF and FCFS. Fig.4 shows the effect of load factor on the response time. The smaller is LF, the lower is the response time. When LF=0.7, the response time of MSA is as much as 30% more than that by LF=0.2. SJN provides the shortest response times amongst all algorithms. The response time of LLF is longest among other algorithms. The value of LF is low, the system is lightly loaded, the price of the sensor provided by sensor resource agent is cheap; sensor user agents with low budget can choose cheap sensor resources to complete jobs under the deadline, so the satisfaction of the sensor user agent is high. When the system is heavily loaded, the price of the sensor resource is expensive; some sensor user agent need more time to complete tasks.

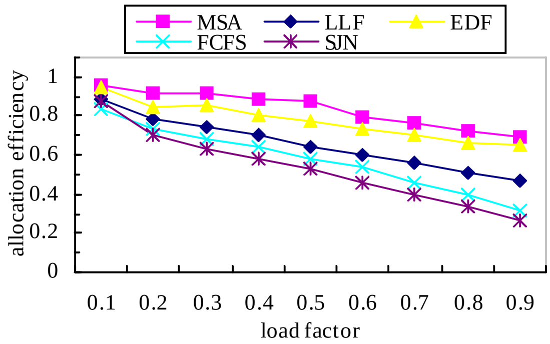

Figure 1: allocation efficiency under various load factor.

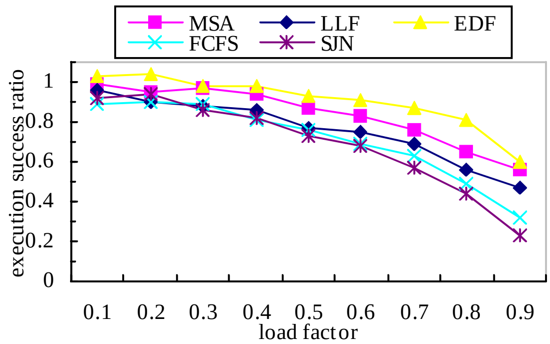

Figure 2: execution success ratio under various load factor.

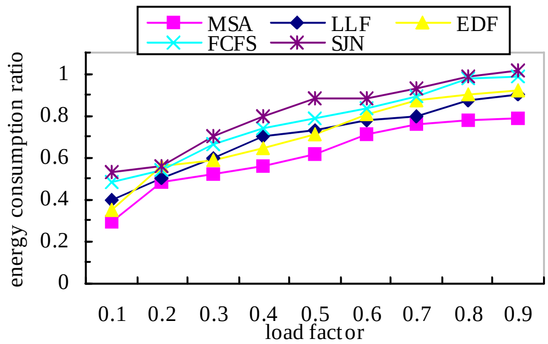

Figure 3: energy consumption ratio under various load factor.

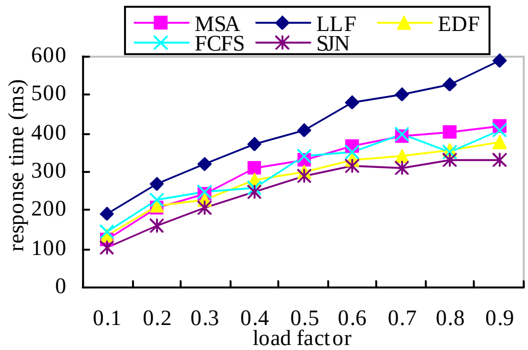

Figure 4: response time under various load factor.

6 Conclusions

The paper presents market based sensor resources allocation in sensor enabled grid computing environment. Since sensor users' tasks might compete for the exclusive usage of the same sensing resource we need to allocate individual sensors to sensor users' tasks. Sensor grid tasks are usually characterized by an uncertain demand for sensing resource capabilities. We model this allocation problem by introducing the sensor utility function. The goal is to find a sensor resource allocation that maximizes the total profit. The paper proposes a distributed optimal sensor resource allocation algorithm. The performance evaluation of proposed algorithm is evaluated and compared with other resource allocation algorithm for sensor grid. In the future, we will consider moving our method to a real grid platform to test its feasibility.

Acknowledgement

The authors thank the editor and the anonymous reviewers for their helpful comments and suggestions. The work was partly supported by the National Natural Science Foundation of China (NSF) under grant (No. 60970064,No.61171075), National Key Basic Research Program of China (973 Program) under Grant No.2011CB302601,Open Fund of the State Key Laboratory of Software Development Environment under Grant (No.SKLSDE-2011KF-01), Program for New Century Excellent Talents in University, China (NCET-08-0806), Fok Ying Tong Education Foundation, China (Grant No. 121067). Any opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of the above agencies.

References

[1] Peter Komisarczuk and Ian Welch, Internet Sensor Grid: Experiences with Passive and Active Instruments, IFIP Advances in Information and Communication Technology, pp. 132-145, 2010

[2] M. Pallikonda Rajasekaran, S. Radhakrishnan, and P. Subbaraj, Sensor grid applications in patient monitoring, Future Generation Computer Systems, 2010, 26(4):569-575

[3] Hock Beng Lim and Danny Lee, An Integrated and Flexible Scheduler for Sensor Grids, UIC 2007, LNCS 4611, pp. 567–578, 2007.

[4] Nikolaos Preve, SEGEDMA: Sensor grid enhancement data management system for Health Care computing, Expert Systems with Applications, 2011,38(3):2371-2380

[5] Huang-Chen Lee, Chun-Yu Lin, Chuan-Yu Cho, Chen-Lung Chan,Yao-Min Fang,Bing-Jean Lee, Chung-Ta King, The design considerations of a sensor grid for monitoring precipitation in debris-flow-prone areas, IPSN '10 Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks, 2010, 362-363

[6] Fox, G.; Ho, A.; Rui Wang; Chu, E.; Isaac Kwan; A Collaborative Sensor Grids Framework ; Collaborative Technologies and Systems, 2008. CTS 2008. International Symposium on, 2008, pp.29 – 38.

[7] Hock Beng Lim; Iqbal, M.; Wenqiang Wang; Yuxia Yao; A Large-Scale Service-Oriented Sensor Grid Infrastructure, Consumer Communications and Networking Conference, 2009. CCNC 2009. 6th IEEE, 2009, pp: 1 – 2

[8] Sanabria, J.; Rivera, W.; Rodriguez, D.; Yunes, Y.; Ross, G.; Sandoval, C.; A Sensor Grid Framework for Acoustic Surveillance Applications, INC, IMS and IDC, 2009. NCM '09. Fifth International Joint Conference on, 2009, pp: 31 - 38

[9] Hock Beng Lim, Mudasser Iqbal, Wenqiang Wang, Yuxia Yao, The National Weather Sensor Grid: a large-scale cyber-sensor infrastructure for environmental monitoring, International Journal of Sensor Networks, 2010,7(1-2):19 - 36

[10] Toshihiro Matsui and Hiroshi Matsuo, A constraint based formalisation for distributed cooperative sensor resource allocation, International Journal of Intelligent Information and Database Systems, 2010, 4(4):307 – 321,

[11] Yan YuJie; Wang Shu; Hao Zhao; MPAS: a Connecting Platform for Integrating Wireless Sensor Network with Grid, Communications, 2005 Asia-Pacific Conference on, 2005 , pp.1000 - 1004

[12] Mohammad Mehedi Hassan, Biao Song and Eui-Nam Huh, A dynamic and fast event matching algorithm for a content-based publish/subscribe information dissemination system in Sensor-Grid, The Journal of Supercomputing, 2010,54(3):330-365

[13] Se-Jin Oh; Chae-Woo Lee; u-Healthcare SensorGrid Gateway for connecting Wireless Sensor Network and Grid Network, Advanced Communication Technology, 2008. ICACT 2008. 10th International Conference on, Volume: 1, 2008, pp.827 - 831

[14] Xiaolin Li; Xinxin Liu; Huanyu Zhao; Nanyan Jiang; Parashar, M.; Autonomic Management of Hybrid Sensor Grid Systems and Applications ; Computer Communications and Networks, 2008. ICCCN '08. Proceedings of 17th International Conference on, 2008, pp:1 – 6.

[15] Toshihiro Matsui, Hiroshi Matsuo, A Formalization for Distributed Cooperative Sensor Resource Allocation, KES-AMSTA 2008, LNAI 4953, pp. 292–301, 2008

[16] Chen-Khong Tham and Rajkumar Buyya, SensorGrid: Integrating Sensor Networks and Grid Computing, CSI Communications, pages 24-29, Vol.29, No.1, Computer Society of India (CSI), Mumbai, India, July 2005

[17] Yong-Kang Ji, Yi Zhang, Zhicheng Xu, and Min-You Wu, Truthful Resource Allocation in Selfish Sensor Web, MSN 2007, LNCS 4864, pp. 683–693, 2007.

[18] Chunlin Li, Layuan Li, Agent Framework to Support Computational Grid, Journal of Systems and Software, 2004, 70(1-2):177-187

[19] Chunlin Li, Layuan Li, Joint QoS Optimization For Layered Computational Grid, Information Sciences, 2007, 177(15):3038-3059

[20] Chunlin Li, Layuan Li, A Distributed Utility-based Two Level Market Solution For Optimal Resource Scheduling In Computational Grid, Parallel Computing, 2005,31(3-4):332-351

[21] Chunlin Li, Layuan Li, Utility based QoS Optimisation Strategy For Multi-Criteria Scheduling on the Grid, Journal of Parallel and Distributed Computing, 2007, 67(2):142-153

[22] JAVASIM, http://javasim.ncl.ac.uk

Analysis of a Single-Agent Search

Aleš Tavčar
Department of Intelligent Systems
Jozef Stefan Institute, Ljubljana, Slovenia
E-mail: ales.tavcar@ijs.si

Keywords: minimin, LRTS, pathology, game tree, single-agent search

Received: February 12, 2012

Game playing is one of the first areas of artificial intelligence that was studied by AI researchers. The developed algorithms and heuristics combined with an ever-increasing computer speed efficiently searched large game trees and thus effectively competed with the best human players. The key advantage comes from the generally accepted notion that a deeper search produces better results. However, while trying to provide a mathematical explanation, researchers have discovered that under certain conditions the opposite situation occurs, i.e., a deeper search results in worse decisions. In this paper we analyse single-agent search algorithms and the influence of three properties of the search trees on the quality of the search. The analysis was performed on one-player search models and in the maze-path-finding problem.

Povzetek: Na modelu enoagentnega preiskovanja in iskanja poti v zemljevidu analiziramo vpliv različnih faktorjev na kvaliteto preiskovanja dveh algoritmov.

1 Introduction

A common way of solving search problems is to represent the problem with a directed search graph. The nodes of the graph correspond to states, e.g., the tiles of a map, and the edges are the moves leading from one state to the other. Moreover, the problem definition includes special states: one initial state and one or several goal states. The solution to the problem is a path leading from the initial to the most appropriate goal node.

Smaller problems can be solved by completely searching the state space [6]. Unfortunately, the search trees of most real-world problems are too large to be searched completely in a reasonable amount of time. Korf introduced an incomplete search method called real-time A* [8] that tackles this problem. Real-time A* expands the game tree to a certain depth, and using a heuristic function it evaluates the quality of the nodes at that depth; it then backs the computed values to the current state. The action leading to the node with the optimal value is then selected. Another algorithm is minimin [8]. Figure 1 demonstrates how the minimin algorithm determines the next action.

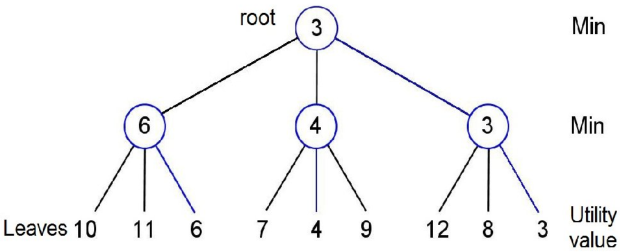

Figure 1: Minimin search.

The nodes of the tree contain utility values, i.e., the quality of a node. For each node the minimum value of its successors is selected and backed-up. Finally, the action leading from the root to the node with the value 3 is selected.

Common sense dictates that a deeper search should provide better decisions, and this can indeed be observed in practice. Evaluating many moves ahead leads to better decisions since obtaining better estimates of the possible strategies can help with selecting the appropriate responses in the current situation. However, while trying to provide a mathematical explanation of this principle, researchers have discovered the opposite: looking ahead does not always provide better decisions. This phenomenon was termed a lookahead pathology [10] and it occurs when the quality of a shallower heuristic search is higher than the quality of a deeper one. This property makes the pathology suitable for measuring the performance of search algorithms at certain depths.

In this paper we study the pathological situations of search algorithms where a deeper search does not guarantee better decisions. On synthetic search-tree models we explore the influence of three important factors on the occurrence of pathological behaviour: the **granularity g** of the heuristic, the **branching factor b** of the game tree and the **similarity s** of the nearby nodes. Next, we perform experimental tests to determine whether these factors have the same influence on the maze-path-finding problem.

This paper is organized as follows. First, we provide a brief literature review of related works in Section 2. In Section 3 we present the game-tree models used during the analysis and we describe the three observed factors. Section 4 explores the influence of these factors in the real-life domain of maze-path finding. Section 5 concludes this article.

2 Related Work

The pathology measure used in this paper was first discovered independently by Nau [10] and Beal [1] in the minimax algorithm [17]. Since then, extensive research and many explanations of the minimax pathology [1,2,7,10,11,12] have broadened our understanding of the principles behind the occurrence of pathological behaviour. Three factors have been identified that greatly contribute to the pathological behaviour in the minimax algorithm: the independence or low similarity of the sibling nodes, the low granularity of the selected heuristic function and the large branching factors of the search trees.

On the other hand, the pathology of a single-agent search has been much less investigated. It was first discovered in 2003 by Bulitko et al. [3], and it was first demonstrated on a synthetic, two-level search tree. Bulitko only demonstrated that the tree is pathological, but he did not provide an adequate explanation. Luštrek [9] continued the research on synthetic search trees and he provided two key factors influencing the pathology. The first of these are certain domain and search-tree properties, with the most important being the difference between the values of the sibling nodes. This is given by the domain and therefore, cannot be altered. The second of these are the factors related to the heuristic function: he showed that consistent and admissible heuristics prevent the pathology. On the other hand, Sadikov and Bratko [14] determined the opposite: pessimistic heuristic functions perform better in the eight-puzzle game. They analysed the influence of several heuristic functions on the number of non-optimal nodes and the quality of the final solutions at greater search depths. Additionally, Bulitko [5] observed that the type of the heuristic function used affects the number of search-tree nodes expanded at greater depths. Their research was limited to the occurrence of pathological behaviour and no insights into the underlying reasons were given.

Piltaver et al. [13] analysed the influence of various properties of the search tree and the heuristic function on the gain and pathology in the eight-puzzle game. They extended the original puzzle to allow additional diagonal moves and thus obtain a different branching factor for the search trees. Next, they analysed the similarity of the sibling nodes and the effect on the pathology. Even though the similarity is given by the domain and therefore cannot be modified at will, they were able to observe that a greater similarity decreases the degree of pathology. Experimentations with heuristic functions included the granularity of the function, the amount of heuristic error and the type of heuristic function (optimistic, pessimistic and uniform).

Nau et al. [12] presented a unifying view of the pathology in the minimax and single-agent searches. The paper shows the interplay between the lookahead pathology and three factors that affect it: the dependence of the sibling nodes in the search tree, the branching factor and the granularity of the heuristic function. The experiments were performed on synthetic trees, the 8-puzzle, two chess endgames and kalah. The content of their paper is related to our work; however, in this paper additional findings relating to the one-player model, as well as an analysis of the maze-path-finding problem, are presented.

3 One-Player Model

In the following section we investigate the influence of three factors on the quality of the search in synthetic search trees: the **granularity g** of the heuristic function (the number of possible heuristic values returned), the **branching factor b** of the search tree (the number of successors of each node), and the **similarity s** of the search tree (the similarity among values of sibling nodes). First, we start with some basic definitions of the model.

3.1 Definitions

Any one- or two-player game can be represented by a game tree in which the nodes are the states and the edges are moves leading from the current to the next state. The goal is to find a path from the initial node to one of the terminal nodes. The quality of a terminal node x is denoted by the utility function u(x). If the overall objective of the search problem is to maximize the cost of reaching a goal, then the player selects the successor with the highest utility score. To each node of the tree a unique value v(n) can be assigned, i.e., the utility value obtained if the player always selects the action leading to the node with the maximum cost:

$$v(n) = \begin{cases} \mathrm{v(n)} & \text{if suc(n)} = \{\} \\ \max_{\mathrm{m \in suc(n)}} c(n,m) + v(m), & \text{otherwise,} \end{cases}$$

where suc(n) is the set of n`s successors. The player following this concept would always select the optimal move. Unfortunately, most search trees are too large to compute v(n) in a reasonable time. Therefore, an approximation of v(n) can be computed using the maximax algorithm, which is the same as the minimin algorithm where the objective is to maximize the gain:

$$v_h(n,d) = \begin{cases} v(n), & \textit{if } suc(n) = \{\} \\ h(n), & \textit{if } d = 0, \\ \max_{m \in suc(n)} c(n,m) + v_h(m,d-1), & \textit{otherwise} \end{cases}$$

where *h(n)* is a *heuristic evaluation function* used to compute an approximation of *v(n)* at a certain *search depth d.* The computed value $v_h(n,d)$ is called a heuristic value, since it is an approximation obtained using a heuristic function. Since the heuristic function is not completely accurate, an error is introduced in the estimation of the utility value. In practice, maximax performs well; therefore, while backing-up values to the root of the search tree it should reduce the error introduced with the heuristic function. Pathology occurs when maximax amplifies the error of the heuristic function. In this paper we will focus on the *decision* *error,* which is the probability of selecting an action that does not lead to a successor with the optimal utility value. The decision error at a certain search depth d and heuristic function h is denoted as *E(h,d)*.

Next, we define the *pathology p* as:

$$p(h,d_1,d_2)=\frac{E(h,d_2)}{E(h,d_1)}, 1 \le d_1 < d_2 \le d_{\max}$$

Pathological behaviour occurs when an error at a deeper level is greater than an error at a shallower level:

$$\exists d_1 < d_2 : \left[E(h,d_1) < E(h,d_2)\right]$$

When this happens the values of $p(h,d_1,d_2) > 1$ indicate that the search algorithm performs poorly, while $p(h,d_1,d_2) < 1$ indicates the absence of pathology, meaning that searching deeper is worthwhile. In order to evaluate whether a problem is pathological we have run several experiments with the Monte Carlo method and compared the averaged *p* with 1.

3.2 Independent game-tree model

This section describes the experiments with synthetic, independent game trees. Namely, we constructed game trees where the sibling nodes are selected randomly and are thus independent of each other.

We start building the game-tree models by assigning utility values to the terminal nodes from a uniform distribution over the interval [0,1]. These values are real; however, we wanted to examine the influence that the number of possible heuristic values has on the search gains. This is called the granularity of the heuristic function. In order to adapt the heuristic function to the desired granularity g, the original values are grouped into g intervals of equal size. Each interval contains values from a certain range from the interval [0,1] and each interval is represented by the mean value of the lower and upper interval bounds. In the one-player model, at small granularities there is a tendency for the values towards the root to converge to a single value: the maximum utility. Due to the nature of the backing-up procedure this phenomenon occurs as soon as every sub tree of the root node contains at least one terminal node with the maximum utility. We solve this by limiting the probability that a maximum utility is reached at the root to at most 50%. In the majority of cases, only half of the values at the root will be the maximum utility. A direct descendant of the root *x* does not have the maximum utility, except when none of the terminal nodes in this sub tree have the maximum utility. Therefore, we have introduced the following equations that describe the relation between the probability of a direct successor of the root having the maximum utility ($P_{\max}(x)$) and the probability of a terminal node having the maximum utility ($P_{\max}(t)$):

$$1-P_{\max}(x)=(1-P_{\max}(t))^{b^{\Delta d}},$$

$$P_{\max}(t)=1-\sqrt[b^{\Delta d}]{1-P_{\max}(x)},$$

where *b* is the branching factor of the tree and Δd is the depth from the node x to the terminal node t. The second equation provides the size of the last interval containing the largest utilities. The boundaries of all the intervals are shifted until the lower boundary of the last interval reaches $1-P_{max}(t)$.

Next, the values from the terminal nodes are propagated throughout the tree using the maximax principle.

3.3 Dependent model

The next task was to introduce a local similarity or a dependence of the sibling nodes into the one-player model.

A search tree has a high dependence or similarity of the sibling nodes when the utility values of the descendants of a node have similar values. In this case, a deeper search is beneficial when the similarity of the sibling nodes increases. The opposite also holds true: the deeper search is becoming less efficient when the similarity of the sibling nodes decreases.

In our model, the dependence is expressed as a parameter $s \in [0,1]$, where $s = 0$ denotes independence between the sibling nodes and $s = 1$ corresponds to the maximum similarity between the sibling nodes. Models with similarity $s = 1$ are constructed by generating b^h random numbers in the interval [0,1] and the values are mapped into granular classes to obtain the desired granularity. Next, the mapped numbers are sorted in increasing order and assigned to the terminal nodes from left to right. The lowest value is assigned to the left-most node, continuing through the inner nodes and finishing with the highest number assigned to the right-most node.

Intermediate similarities for the models ($0 < s < 1$) are created by randomly mixing the terminal nodes from the independent and completely dependent trees. All the sibling terminal nodes in the independent tree are replaced with terminal nodes from the dependent tree with the probability *s*.

The introduced measure of similarity is useful when we are dealing with game-tree models; however, local similarity in an arbitrary game is expressed in a different way. If we want to compare results from the models and the game trees we need a common similarity measure. For this purpose the *clustering factor f* was used. Roughly, it is the ratio between the standard deviation of the utility values of the sibling nodes averaged across all the possible parent nodes, and the standard deviation of the utilities in the entire search tree [15]. When *f* is low, then the sibling nodes are similar, and vice versa. We calculated the clustering factor for every similarity *s*.

3.4 Evaluation

This section describes experiments with the one-player models. The performance of the maximax algorithm was estimated by observing how the degree of pathology is influenced by three factors: the branching factor *b*, the granularity *g*, and the similarity *s*. For each combination of values for the three observed factors 10,000 synthetic trees were generated using the one-player model. The following values were systematically used for each of the three factors: b = 2,3,…,10; g = 2,3,…,300; and s = 0.0,0.1,…, 1.0.

In the first experiment we observed how the quality of the search is influenced by the branching factor and the granularity in independent trees. Figure 2 shows the pathology p in relation to the granularity g and the similarity s in the results of the experiment. On one hand the search is inefficient at lower granularities and as the branching factor increases. On the other hand, a deeper search becomes worthwhile with an increased granularity and at lower branching factors.

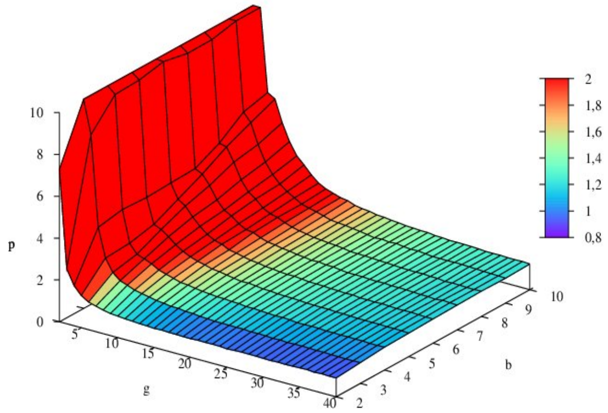

Figure 2: The degree of pathology for independent search trees.

In the next experiment we computed the required granularity to avoid situations where a deeper search is inefficient, as a function of the branching factor b and the local similarity s. The surface in Figure 3 represents the border between pathological and non-pathological behaviour where p = 1. The area below the surface is pathological, while the area above the surface is non-pathological. The results are as expected: higher similarity decreases the granularity needed to avoid search inefficiency, while a higher branching factor increases the needed granularity.

The unified influence of the branching factor, the local similarity and the granularity on the quality of the search is presented in Figure 4. Again, the pathology was used to estimate the performance of the maximax algorithm. Darker dots denote a slightly weak performance ($1<=p<1.5$) and an extremely weak performance ($p>1.5$), respectively. The dots in light grey denote the parameter settings where a deeper search is worthwhile.

The observed relations among the factors are the same as presented in the previous figures. The degree of pathology decreases with the increased granularity g and the local similarity s and increases with the branching factor b. Moreover, one can see that the similarity of the sibling nodes affects the degree of pathology the most. In a high-similarity search tree the influence of the granularity and the branching factor on the pathology is diminished. This can be clearly observed when comparing independent and completely dependent trees. The pathological behaviour is present for most granularities and branching factors in the independent trees, while the pathology is mostly eliminated in the completely dependent trees. In addition, the degree of pathology is not noticeably increasing with higher branching factors. Another effect that can be observed in Figure 4 is that large values of b decrease the difference in the degree of pathology between values near the pathological and non-pathological border. For example, at lower branching factors, values transit from a high pathology degree ($p >= 1.5$) to a low pathology degree ($p < 0.5$) relatively quickly.

The next experiment investigated how well the synthetic trees model the 8-puzzle, a real-world game.

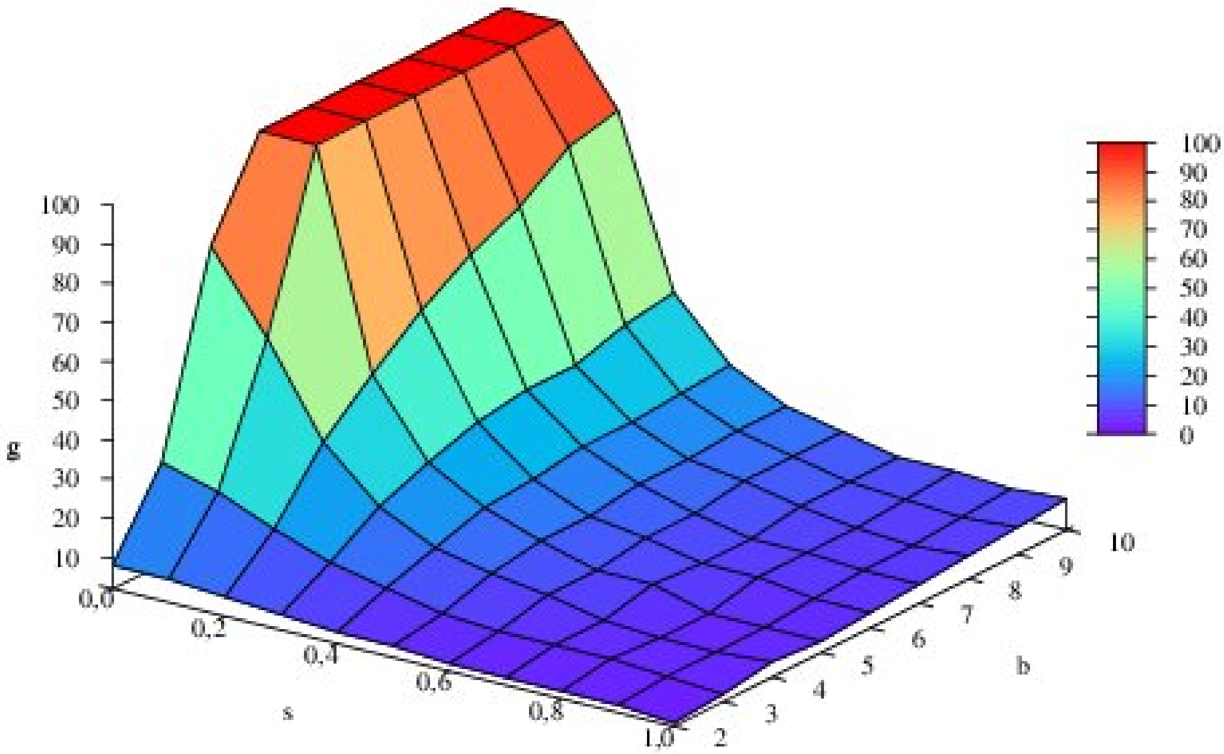

Figure 3: Granularity values needed to avoid pathology in a deeper search.

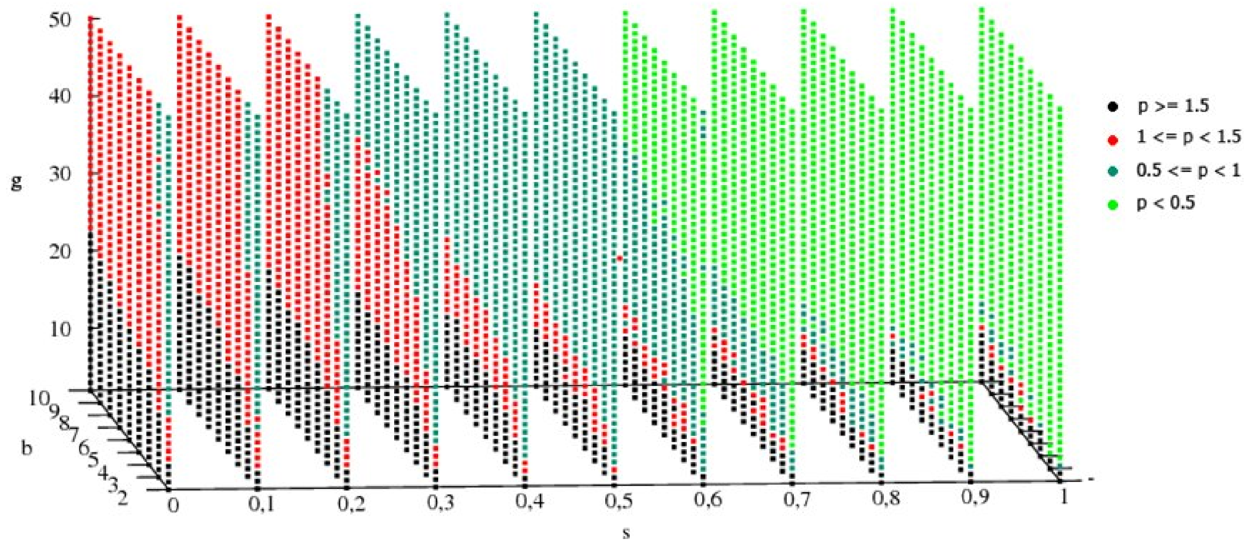

Figure 4: Degree of pathology in the minimin search model.

From [13] we obtained the degree of pathology for two heuristic functions, with a branching factor of 2 and a similarity of 0.4. The similarity in the 8-puzzle was measured with the clustering factor; therefore, we had to map the value of the clustering factor to the similarity used in the models as described in Subsection 3.3. The comparison between the model and the 8-puzzle is shown in Figure 5. Qualitatively, the performance of the model is similar to the 8-puzzle. Quantitatively, the minimax algorithm from the model performs worse at lower granularities (g < 20), but performs very similarly at higher granularities. The granularity needed to avoid inefficiency in a deeper search is higher in the model than in the 8-puzzle. The reason that the synthetic model performs worse than the Manhattan distance is probably related to the specific properties of the function, e.g., the Manhattan function is exact for distances less than eight moves to the finish.

4 Pathology in the Maze Search

This section presents experimental tests on a real-world problem, i.e., maze path finding, and an analysis of the Learning Real-Time Search (LRTS) algorithm [4] The algorithm conducts a lookahead search centred on the current state and generates all the states up to d moves away from the current state.

The analysis of LRTS consisted of determining if the algorithm selects the optimal first move from the initial node. First, we computed the shortest path from the initial to the goal node of the maze to determine the correct move from the initial node. Next, we run one basic step of the LRTS algorithm with different lookahead depths (d=1 and d=5). Finally, we compared the move computed with the two LRTS runs with the computed optimal move. If the move selected with LRTS differs from the optimal then this is a decision error for a certain lookahead depth.

As in the single-player model, the measure of the search quality is the pathology p. In addition we compare the performance of the LRTS algorithm and the minimin algorithm used in the model.

Using the Randomized Kruskal's algorithm we created mazes of different sizes and structures. In the generated mazes it is possible to vary only the granularity g of the utility and the heuristic function since in real-world problems the branching factor b and the local similarity s are part of the problem. However, we were still able to generate mazes with different local similarities and calculate the corresponding degree of pathology. In Figure 6 it is clear that the degree of pathology decreases with the similarity of the sibling nodes. The similarity is measured with the clustering factor; therefore, a smaller clustering factor means a larger similarity. The similarities in the figure range from 0.03 to 0.67, i.e., game trees that are almost completely dependent to games trees with a smaller dependence. The same cannot be achieved for the branching factor since the average branching factor in an arbitrary maze is 2.

The experiments thus analyse the influence of the amount of granularity and type of heuristic function on the performance of the LRTS algorithm. The first step of the experiments is to compute the utilities for all the positions of the maze. The utility of a given position is equal to the minimum number of moves needed to reach

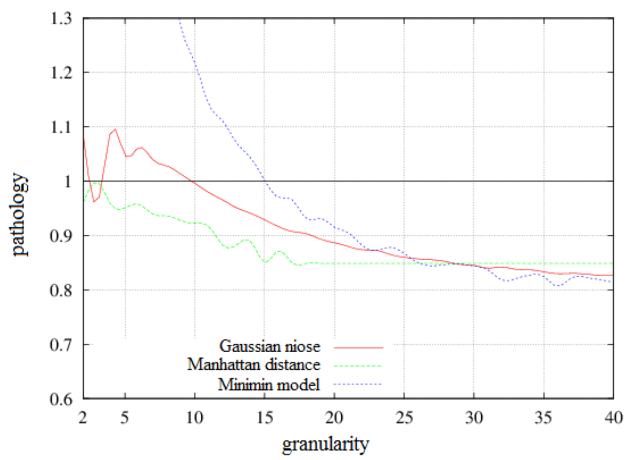

Figure 5: Search efficiency of three single-agent search problems.

the exit of the maze from that position. Using the retrograde analysis [16] we start at the goal position and while moving towards the start we assign to each position the number of moves currently traversed. Next, the heuristic values are computed from the utility values using the *generic heuristic evaluation function h*. The error of the heuristic function is the difference between the utility and the heuristic values and is modelled using Gaussian noise. The advantage of such an approach is the ability to adjust the noise distribution and thus approximate any heuristic function used in practice. In our experiments the heuristic values were computed by adding the Gaussian noise with a standard deviation σ = 2.5 to the utility values. The selected standard deviation is the same as the standard deviation of the Manhattan-distance heuristic function.

The desired granularity *g* of the utility and the heuristic values is obtained by creating *g* intervals of equal size. We start by limiting the heuristic values to the [0,N] interval, where $N = \max_n\{h^*(n)\}+\lfloor\sigma\rfloor+1$. If any values are lower than 0, it is set to 0, and if it is greater than N, it is set to N. Next, all the heuristic values are multiplied by g/N to further reduce the interval of possible heuristic values to [0,g].

The second property of the heuristic function that we investigated is how the heuristic error is applied to the utility values. We considered the influence of three heuristic functions: balanced, pessimistic and optimistic heuristic functions. Optimistic heuristic functions always underestimate the utility value, i.e., the heuristic estimate is smaller than the utility value. On the other hand, pessimistic heuristic functions always overestimate the utility value, and balanced heuristic functions are a combination of both. Optimistic and pessimistic heuristic functions were obtained by subtracting or adding the absolute value of the Gaussian noise to the utility values.

Pessimistic functions were found to be less prone to a lookahead pathology [14]; therefore, the use of this heuristic function should improve the performance of the LRTS algorithm.

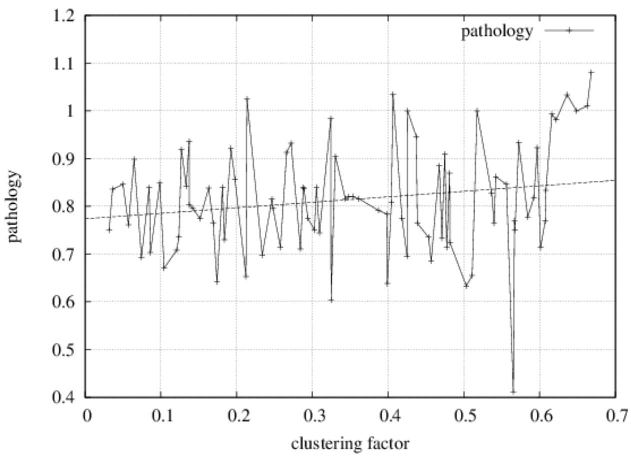

Figure 6: The influence of the similarity on the degree of pathology.

The last step is to analyse the influence of the selected factors on the performance of the algorithm. Using Monte Carlo simulations we generated 10,000 mazes for different granularities and heuristic functions. The results of the analysis are presented in Figure 7. The figure shows how the three heuristic functions and the granularity contribute to the LRTS algorithm`s quality of search. For all three functions it is clear that an increased granularity increases the quality of the search, i.e., decreases the degree of pathology. Moreover, the obtained results confirm that pessimistic heuristic functions perform better than optimistic ones. At granularities higher than 5, a deeper search is beneficial and the degree of pathology stabilizes around 0.75 at g>13. The performance of the balanced, generic, heuristic function is worse than that of the pessimistic

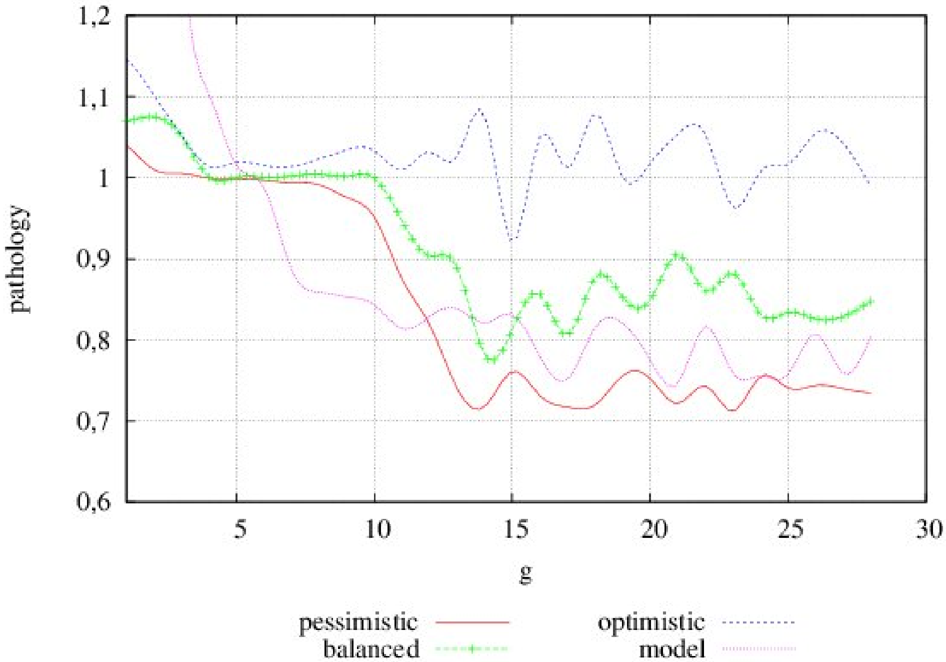

Figure 7: Performance of the three heuristic functions and the reference model.

heuristic function, but better than that of the optimistic heuristic function. For comparison, we added the performance of the minimin algorithm from the one-player model with the appropriate local similarity. The model behaves in a similar way to the LRTS algorithm with the pessimistic heuristic function.

5 Discussion

In this paper we analysed the performance of two one-player search algorithms, i.e., maximax and LRTS, under particularly demanding conditions. The analysis was performed on a model for one-player games and in the maze-path-finding problem. Three factors were considered during the analysis of the quality of the search for both algorithms: the granularity of the heuristic function, the local similarity of the sibling nodes and the branching factor of the game tree. We were unable to study the influence of the last factor in the maze domain since the branching factor is usually fixed in real-world domains.

Based on the experimental results several observations can be made. First, an increased granularity of the heuristic function increases the search gain of both algorithms. Second, in the one-player model an increased branching factor of the game tree reduces the search gain. Third, pessimistic heuristic functions have higher search gains than optimistic ones in the maze domain. Finally, the similarity of the sibling nodes is the single most important factor to improve the search gain: a higher similarity improves the search results.

We confirmed that under the specific, demanding conditions, both search algorithms perform poorly: a deeper search in these conditions produces worse results than a shallow search. The results are consistent with recent publications, but a certain amount of novelty is demonstrated by the analysis of the search-path finding and analyses of the LRTS algorithm.

References

[1] Beal, D. F. (1980). *An Analysis of Minimax*. In Advances in Computer Chess 2, M.R.B. Clarke (Ed.), pp. 103-109.

[2] Bratko, I., Gams, M. (1982). *Error analysis of the minimax principle*, In Advances in Computer Chess 3 M. Clarke (Ed.), Pergamon Press, pp. 1-15.

[3] Bulitko, V. (2003). *Lookahead Pathologies and Meta-level Control in Real-time Heuristic Search*. Proceedings of 15th Euromicro Conference on Real-Time Systems, Porto, Portugal, pp. 13-16.

[4] Bulitko, V., Lee, G. (2006). *Learning in real time search:A unifying framework*. *JAIR* 25:119–157.

[5] Bulitko, V., Li, L., Greiner, R. and Levner, I. (2003). *Lookahead Pathologies for Single Agent Search*. Proceedings of International Joint Conference on Artificial Intelligence (IJCAI), Acapulco, Mexico, pp. 1531-1533.

[6] Hart, P.E., Nilsson, N.J. and Raphael, B. (1968). *A Formal Basis for the Heuristic Determination of Minimum Cost Paths*. IEEE Transactions on Systems Science and Cybernetics 4(2), pp. 100-107.

[7] Kaluža, B., Luštrek, M., Gams, M., Tavčar, A. (2007). *Pathology in Minimax Searching.* Proceedings of the 16th International Electrotechnical and Computer Science Conference, pp. 111-114.

[8] Korf, R. E. (1990). *Real-Time Heuristic Search*. Artificial Intelligence 42(2, 3), pp. 189-211.

[9] Luštrek, M. (2005). *Pathology in Single-agent Search.* Proceedings of the 8th International Multiconference Information Society, Slovenia, Ljubljana, pp. 345-348.

[10] Luštrek, M., Gams, M., Bratko, I. (2006). *Is real-valued minimax pathological*. Artifcial Intelligence 170 (6-7), pp. 620-642.

[11] Nau, D. S. (1979). *Quality of Decision versus Depth of Search on Game Trees*. Ph. D. thesis, Duke University.

[12] Nau, D. S., Luštrek, M., Parker, A., Bratko, I., and Gams, M. (2010). *When is it Better not to Look Ahead?*, Artificial Intelligence, 174, pp. 1323-1338.

[13] Piltaver, R., Luštrek, M., Gams, M. *The pathology of heuristic search in the 8-puzzle.* Journal of Experimental & Theoretical Artificial Intelligence, DOI:10.1080/0952813X.2010.545997.

[14] Sadikov, A., Bratko, I. (2006). *Pessimistic Heuristics Beat Optimistic Ones in Real-time Search.* Proceedings of the European Conference on Artificial Intelligence, Italy, Riva del Garda, pp. 148-152.

[15] Sadikov, A., Bratko, I., Kononenko, I. (2005). *Bias and pathology in minimax search.* Theoretical Computer Science 349(2), pp. 261-281.

[16] Thompson, K. (1986). *Retrograde Analysis of Certain Endgames*. ICCA Journal,3, pp. 131-139.

[17] Von Neumann, J. (1928). *Zur Theorie der Gesellschaftsspiele*. Mathematische Annalen, 100, pp. 295-320.

Graph Theory Algorithms for Mobile Ad Hoc Networks

Natarajan Meghanathan
Department of Computer Science, Jackson State University
Jackson, MS 39217, USA
E-mail: natarajan.meghanathan@jsums.edu

Keywords: mobile ad hoc networks, routing protocols, graph algorithms, unicast, multi-path, multicast and broadcast

Received: April 22, 2011

Various simulators (e.g., ns-2 and GloMoSim) are available to implement and study the behavior of the routing protocols for mobile ad hoc networks (MANETs). But, students and investigators who are new to this area often get perplexed in the complexity of these simulators and lose the focus in designing and analyzing the characteristics of the network and the protocol. Most of the time would be spent in learning the existing code modules of the simulator and the logical flow between the different code modules. The purpose of this paper would be to illustrate the applications of Graph Theory algorithms to study, analyze and simulate the behavior of routing protocols for MANETs. Specifically, we focus on the applications of Graph Theory algorithms to determine paths, trees and connected dominating sets for simulating and analyzing respectively unicast (single-path and multi-path), multicast and broadcast communication in mobile ad hoc networks (MANETs). We will discuss the (i) Dijkstra's shortest path algorithm and its modifications for finding stable paths and bottleneck paths; (ii) Prim's minimum spanning tree algorithm and its modification for finding all pairs smallest and largest bottleneck paths; (iii) Minimum Steiner tree algorithm to connect a source node to all the receivers of a multicast group; (iv) A node-degree based algorithm to construct an approximate minimum connected dominating set (CDS) for sending information from one node to all other nodes in the network; and (v) Algorithms to find a sequence of link-disjoint, node-disjoint and zone-disjoint multi-path routes in MANETs.

Povzetek: Prispevek opisuje algoritme za mobilna omrežja.

1 Introduction

A Mobile Ad hoc Network (MANET) is a dynamically changing infrastructureless and resource-constrained network of wireless nodes that may move arbitrarily, independent of each other. The transmission range of the wireless nodes is often limited, necessitating multi-hop routing to be a common phenomenon for communication between any two nodes in a MANET. Various routing protocols for unicast, multicast, multi-path and broadcast communication have been proposed for MANETs. The communication structures that are often determined include: a path (for unicast – single-path and multi-path routing), a tree (for multicast routing) and a connected dominating set – CDS (for broadcast routing). Within a particular class, it is almost impossible to find a single routing protocol that yields an optimal communication structure with respect to different route selection metrics and operating conditions.

Various simulators such as ns-2 [5] and GloMoSim [20] are available to implement and study the behavior of the routing protocols. But, students and investigators who are new to this area often get perplexed in the complexity of these simulators and lose the focus in designing and analyzing the characteristics of the network and the protocol. Most of the time would be spent in learning the existing code modules of the simulator and the logical flow between the different code modules. The purpose of this paper would be to illustrate the applications of Graph Theory algorithms to study, analyze and simulate the behavior of routing protocols for MANETs. We will discuss the applications of Graph Theory algorithms for unicast (single-path and multi-path), multicast and broadcast communication in MANETs.

An ad hoc network is often approximated as a unit disk graph [10]. In this graph, the vertices represent the wireless nodes and an edge exists between two vertices u and v if the normalized Euclidean distance (i.e., the physical Euclidean distance divided by the transmission range) between u and v is at most 1. Two nodes can communicate only if each node lies within (or on the edge of) the unit disk of the other node. The unit disk graph model neatly captures the behavior of many practical ad hoc networks and would be used in the rest of this paper for discussing the algorithms to simulate the MANET routing protocols.

Most of the contemporary routing protocols proposed in the MANET literature adopt a Least Overhead Routing Approach (LORA) according to which a communication structure (route, tree or CDS) discovered through a global flooding procedure would be used as long as the communication structure exist, irrespective of the structure becoming sub-optimal since the time of its discovery in the MANET. We will also adopt a similar strategy and focus only on discovering a communication structure on a particular network graph taken as a snapshot during the functioning of the MANET. Such a graph snapshot would be hereafter referred to as a 'Static Graph' and a sequence of such static graphs over the duration of the MANET simulation session would be called a 'Mobile Graph'. A communication structure determined on a particular static graph would be then validated for its existence in the subsequent static graphs and once the structure breaks, the appropriate graph algorithm can be invoked on the static graph corresponding to that particular time instant and the above procedure would be continued for the rest of the static graphs in the mobile graph. We use the big-O notation to express the theoretical worst-case run-time complexity of the algorithms discussed in this paper. Given a problem size x, where x is usually the number of items, we say $f(x) = O(g(x))$, when there exists positive constants c and k such that $0 \leq f(x) \leq cg(x)$, for all $x \geq k$ [4].

The rest of this paper is organized as follows: Section 2 reviews related work on unicast, multicast, broadcast and multi-path communication in MANETs. In the subsequent sections, we discuss graph theory algorithms for unicast communication (Section 3), the tree-based algorithms for multicast communication (Section 4), a maximum density-based CDS algorithm for broadcast communication (Section 5) and multi-path algorithms for determining link-disjoint, node-disjoint and zone-disjoint routes (Section 6) in MANETs. Section 7 concludes the paper and Section 8 discuss future research directions in this area. Throughout the paper, the terms 'route' and 'path', 'link' and 'edge', 'message' and 'packet' are used interchangeably. They mean the same.

2 Background Work

2.1 Unicast Communication in MANETs

There are two broad classifications of unicast routing protocols: minimum-weight based routing and stability-based routing. Routing protocols under the minimum-weight category have been primarily designed to optimize the hop count of source-destination (s-d) routes. Some of the well-known minimum-hop based routing protocols include the Dynamic Source Routing (DSR) protocol [8] and the Ad hoc On-demand Distance Vector (AODV) routing protocol [16]. The stability-based routing protocols aim to minimize the number of route failures and in turn reduce the number of flooding-based route discoveries. Some of the well-known stability-based routing protocols include the Flow-Oriented Routing Protocol [18] and the Node Velocity-based Stable Path (NVSP) routing protocol [12]. In [13] and [14], it was observed that there exists a stability-hop count tradeoff and it is not possible to simultaneously optimize both the hop count as well as the number of route discoveries.

The DSR protocol is a source routing protocol that requires the entire route information to be included in the header of every data packet. However, because of this feature, intermediate nodes do not need to store up-to-date routing information in their routing tables. Route discovery is by means of the broadcast query-reply cycle. The Route Request (RREQ) packet reaching a node contains the list of intermediate nodes through which it has propagated from the source node. After receiving the first RREQ packet, the destination node waits for a short time period for any more RREQ packets, then chooses a path with the minimum hop count and sends a Route Reply (RREP) along the selected path. Later, if any new RREQ is received through a path with hop count less than that of the selected path, another RREP would be sent on the latest minimum hop path discovered.

The AODV protocol, like DSR, is also a shortest path based routing protocol. However, it is table-driven. Upon receiving an unseen RREQ packet (with the highest sequence number seen so far), an intermediate node records the upstream node (sender) of the RREQ packet in its routing table entry for the source-destination route. The intermediate node then forwards the RREQ packet by incrementing the hop count of the path from the source node. The destination node receives RREQ packets on several routes and selects that RREQ packet that traversed on the minimum-hop path to the destination node. The RREP packet is then sent on the reverse of this minimum-hop path towards the source node. The destination node includes the upstream node from which the RREQ was received as the downstream node on the path from the destination node to the source node. An intermediate node upon receiving the RREP packet will check whether it has been listed as the downstream node ID. In that case, the intermediate node processes the RREP packet and completes its routing table by including the sender of the RREP packet as the next hop node on the path from the source node towards the destination node. The intermediate node then replaces its own ID in the RREP downstream node entry with the ID of the upstream node that it has in its routing table for the path from the source node to the destination node.

The FORP protocol has been observed to discover the sequence of most stable routes among the contemporary stable path routing protocols [13]. FORP utilizes the mobility and location information of the nodes to approximately predict the expiration time (LET) of a wireless link. The minimum of LET values of all wireless links on a path is termed as the route expiration time (RET). The route with the maximum RET value is selected as the desired route. Each node is assumed to be able to predict the LET values of the links with its neighboring nodes based on the information regarding the current position of the nodes, velocity, the direction of movement, and transmission range. FORP assumes the availability of location-update mechanisms like Global Positioning System (GPS) [6] to identify the location of the nodes and also requires each node to periodically broadcast its location and mobility information to its neighbors through beacons.

The NVSP protocol is the only beaconless routing protocol that can discover long-living stable routes without significant increase in the hop count per path.

FORP discovers routes that have a significantly larger hop count than the minimum value. NVSP only requires each intermediate node to include its velocity in the RREQ packets propagated via flooding from the source node to the destination node. With flooding, each intermediate node forwards the RREQ packet exactly once, the first time the node sees the packet as part of a particular route discovery session. The destination node receives the RREQ packets through several paths and determines the bottleneck velocity of each of those paths. The bottleneck velocity of a path is the maximum among the velocities of the intermediate nodes on the path. The destination node chooses the path with the minimum bottleneck velocity and sends a RREP packet along that path. In case of a tie, the destination node chooses the path with the lowest hop count and if the tie could not be still broken, the destination node chooses an arbitrary path among the contending paths.

2.2 Multicast Communication in MANETs

The Multicast communication refers to sending messages from one source node to a set of receiver nodes in a network. The receiver nodes form the multicast group and we typically find a tree that connects the source node to the multicast group members such that there is exactly one path from the source node to each receiver node. The tree could be constructed based on either one of the following two objectives: (i) Shortest path tree – the tree would have the minimum hop count paths from the source node to each receiver node and (ii) Steiner tree – the tree would have the minimum number of links spanning the source node and the multicast group members. Both these trees cannot be simultaneously built and there would always be a tradeoff between the above two objectives [14]. The Multicast Extension of the Ad hoc On-demand Distance Vector (MAODV) protocol and the Bandwidth Efficient Multicast Routing Protocol (BEMRP) are respectively examples of the minimum hop and minimum link based multicast protocols.

MAODV [15] is the multicast extension of the AODV unicast routing protocol. Here, a receiver node joins the multicast tree through a member node that lies on the minimum-hop path to the source node. A potential receiver node wishing to join the multicast group broadcasts a RREQ message. If a node receives the RREQ message and is not part of the multicast tree, the node broadcasts the message in its neighborhood and also establishes the reverse path by storing the state information consisting of the group address, requesting node id and the sender node id in a temporary cache. If a node receiving the RREQ message is a member of the multicast tree and has not seen the RREQ message earlier, the node waits to receive several RREQ messages and sends back a RREP message on the shortest path to the receiver node. The member node also informs in the RREP message, the number of hops from itself to the source node. The potential receiver node receives several RREP messages and selects the member node which lies on the shortest path to the source node. The receiver node sends a Multicast Activation (MACT) message to the selected member node along the chosen route. The route from the source node to the receiver node is set up when the member node and all the intermediate nodes in the chosen path update their multicast table with state information from the temporary cache.

According to BEMRP [17], a newly joining node to the multicast group opts for the nearest forwarding node in the existing tree, rather than choosing a minimum-hop count path from the source node of the multicast group. As a result, the number of links in the multicast tree is reduced leading to savings in the network bandwidth. Multicast tree construction is receiver-initiated. When a node wishes to join the multicast group as a receiver node, it initiates the flooding of Join control packets targeted towards the nodes that are currently members of the multicast tree. On receiving the first Join control packet, the member node waits for a certain time before sending a Reply packet. The member node sends a Reply packet on the path, traversed by the Join control packet, with the minimum number of intermediate forwarding nodes. The newly joining receiver node collects the Reply packets from different member nodes and would send a Reserve packet on the path that has the minimum number of forwarding nodes from the member node to itself.

2.3 Broadcast Communication in MANETs

Broadcast communication refers to sending a message from one node to all the other nodes in the network. Since MANET topology is not fully connected as nodes operate with a limited transmission range, multi-hop communication is a common phenomenon in routing. As a result, a message has to be broadcast by more than one node (in its neighborhood) so that the message can reach all the nodes in the network. An extreme case of broadcasting is called flooding wherein each node broadcasts the message among its neighbors, exactly once, when the message is seen for the first time. This ensures that the message is received by all the nodes in the network. However, flooding would cause unnecessary retransmissions, exhausting the network bandwidth and the energy reserves at the nodes.

Connected Dominating Sets (CDS) are considered to be very efficient for broadcasting a message from one node to all the nodes in the network. A CDS is a sub graph of a given undirected connected graph such that all nodes in the graph are included in the CDS or directly attached to a node (i.e., covered by a node) in the CDS. A Minimum Connected Dominating Set (MCDS) is the smallest CDS (in terms of the number of nodes in the CDS) for the entire graph. For a virtual backbone-based route discovery, the smaller the size of the CDS, the smaller is the number of unnecessary retransmissions. If the RREQ packets of a broadcast route discovery process get forwarded only by the nodes in the MCDS, we will have the minimum number of retransmissions. Unfortunately, the problem of determining the MCDS in an undirected graph, like that of the unit disk graph considered for modeling MANETs, is NP-complete. In

[1], [2] and [3], efficient algorithms have been proposed to approximate the MCDS for wireless ad hoc networks. A common thread among these algorithms is to give preference to nodes with high neighborhood density (i.e., a larger number of uncovered neighbors) for inclusion in the MCDS.

2.4 Multi-path Communication in MANETs

MANET routing protocols incur high route discovery latency and also incur frequent route discoveries in the presence of a dynamically changing topology. Recent research has started to focus on multi-path routing protocols for fault tolerance and load balancing. Multi-path on-demand routing protocols tend to compute multiple paths, at both the traffic sources as well as at intermediary nodes, in a single route discovery attempt. This reduces both the route discovery latency and the control overhead as a route discovery is needed only when all the discovered paths fail. Spreading the traffic along several routes could alleviate congestion and bottlenecks. Multi-path routing also provides a higher aggregate bandwidth and effective load balancing as the data forwarding load can be distributed over all the paths.

Multi-paths can be of three types: link-disjoint, node-disjoint and zone-disjoint. For a given source node s and destination node d, the set of link-disjoint s-d routes comprises of paths that have no link present in more than one constituent s-d path. Similarly, the set of node-disjoint s-d routes comprises of paths that have no node (other than the source node and destination node) present in more than one constituent s-d path. A set of zone-disjoint s-d routes comprises of paths such that an intermediate node in one path is not a neighbor node of an intermediate node in another path. Multi-path on-demand routing protocols tend to compute multiple paths between a source-destination (s-d) pair, in a single route discovery attempt. A new network-wide route discovery operation is initiated only when all the s-d paths fail. The Split Multi-path Routing (SMR) protocol [11], the AODV-Multi-path (AODVM) protocol [19] and the Zone-Disjoint multi-path extension to the DSR (ZD-DSR) protocol [7] are respectively well-known examples for link-disjoint, node-disjoint and zone-disjoint multi-path routing protocols.

In SMR, the intermediate nodes forward RREQs that are received along a different link and with a hop count not larger than the first received RREQ. The destination node selects the route on which it received the first RREQ packet (which will be a shortest delay path), and then waits to receive more RREQs. The destination node then selects the path which is maximally disjoint from the shortest delay path. If more than one maximally disjoint path exists, the tie is broken by choosing the path with the shortest hop count.

In AODVM, an intermediate node does not discard duplicate RREQ packets and records them in a RREQ table. The destination node responds with an RREP for each RREQ packet received. An intermediate node, on receiving the RREP, checks its RREQ table and forwards the packet to the neighbor that lies on the shortest path to the source node. The neighbor entry is then removed from the RREQ table. Also, whenever a node hears a neighbor node forwarding the RREP packet, the node removes the entry for the neighbor node in its RREQ table.

The Zone-Disjoint Multi-path extension of the Dynamic Source Routing (ZD-MPDSR) protocol proposed for an omni-directional system works as follows: Whenever a source node has no route to send data to a destination node, the source node initiates broadcast of the RREQ messages. The number of active neighbors for a node indicates the number of neighbor nodes that have received and forwarded the RREQ message during a route discovery process. The RREQ message has an ActiveNeighborCount field and it is updated by each intermediate node before broadcasting the message in the neighborhood. When an intermediate node receives the RREQ message, it broadcasts a 1-hop RREQ-query message in its neighborhood to determine the number of neighbors who have also seen the RREQ message. The number of RREQ-query-replies received from the nodes in the neighborhood is the value of the ActiveNeighborCount field updated by a node in the RREQ message. The destination node receives several RREQ messages and selects the node-disjoint paths with lower ActiveNeighborCount values and sends the RREP messages to the source node along these paths. Even though the selection of the zone-disjoint paths with lower number of active neighbors will lead to reduction in the end-to-end delay per data packet, the route acquisition phase will incur a significantly longer delay as RREQ-query messages are broadcast at every hop (in addition to the regular RREQ message) and the intermediate nodes have to wait to receive the RREQ-query and reply messages from their neighbors. This will significantly increase the control overhead in the network.

3 Graph Theory Algorithms for Unicast Communication in MANETs

In a graph theoretic context, we illustrate that the minimum-weight (minimum-hop) based routing protocols could be simulated by running the shortest-path Dijkstra algorithm [4] on a mobile graph (i.e. a sequence of static graphs). We then illustrate that the NVSP and FORP protocols could be simulated by respectively solving the smallest bottleneck and the largest bottleneck path problems – each of which could be implemented as a slight variation of the shortest path Dijkstra algorithm. In addition, we also illustrate that the Prim's minimum spanning tree algorithm and its modification to compute the maximum spanning tree can be respectively used to determine the 'All Pairs Smallest Bottleneck Paths' and 'All Pairs Largest Bottleneck Paths' in a weighted network graph.

Informatica **36** (2012) 185–200

3.1 Shortest Path Problem

Given a weighted graph G = (V, E), where V is the set of vertices and E is the set of weighted edges, the shortest path problem is to determine a minimum-weight path between any two nodes (identified as source node s and destination node d) in the graph. The execution of the Dijkstra algorithm (pseudo code in Figure 1) on a weighted graph starting at the source node s results in a shortest path tree rooted at s. In other words, the Dijkstra algorithm will actually return the minimum-weight paths from the source vertex s to every other vertex in the weighted graph. If all the edge weights are 1, then the minimum-weight paths are nothing but minimum-hop paths.

Begin Algorithm *Dijkstra-Shortest-Path* (*G*, *s)*
For each vertex $v \in V$
weight [*v*] ← ∞ // an estimate of the minimum-weight path from *s* to *v*
End For
weight [*s*] ← 0
$S \leftarrow \Phi$ // set of nodes for which we know the minimum-weight path from *s*
$Q \leftarrow V$ // set of nodes for which we know estimate of the minimum-weight path from *s*
While $Q \neq \Phi$
u ← EXTRACT-MIN (*Q*)
$S \leftarrow S \cup \{u\}$
For each vertex *v* such that $(u, v) \in E$
If *weight* [*v*] > *weight* [*u*] + *weight* (*u*, *v*) **then**
weight [*v*] ← *weight* [*u*] + *weight* (*u*, *v*)
Predecessor (*v*) ← *u*
End If
End For
End While
End *Dijkstra-Shortest-Path*

Figure 1: Pseudo Code for Dijkstra's Shortest Path Algorithm.

Dijkstra algorithm proceeds in iterations. To begin with, the weights of the minimum-weight paths from the source vertex to every other vertex is assumed to be +∞ (as estimate value, indicating that the paths are actually not known) and from the source vertex to itself is assumed to be 0. During each iteration, we determine the shortest path from the source vertex *s* to a particular vertex *u*, which would be the vertex with the minimum weight among the vertices that have been not yet optimized (i.e. for which the shortest path has not been yet determined). We then explore the neighbors of *u* and determine whether we can reach any of the neighbor vertices, say *v*, from *s* through *u* on a path with weight less than the estimated weight of the current path we know from *s* to *v*. If we could find such a neighbor *v*, then we set the predecessor of *v* to be vertex *u* on the shortest path from *s* to *v*. This step is called the relaxation step and is repeated over all iterations. The darkened edges shown in the working example of Figure 2 are the edges that are part of the shortest-path tree rooted at the source vertex *s*. The run-time complexity of the Dijkstra's shortest path algorithm is $O(|V|^2)$.

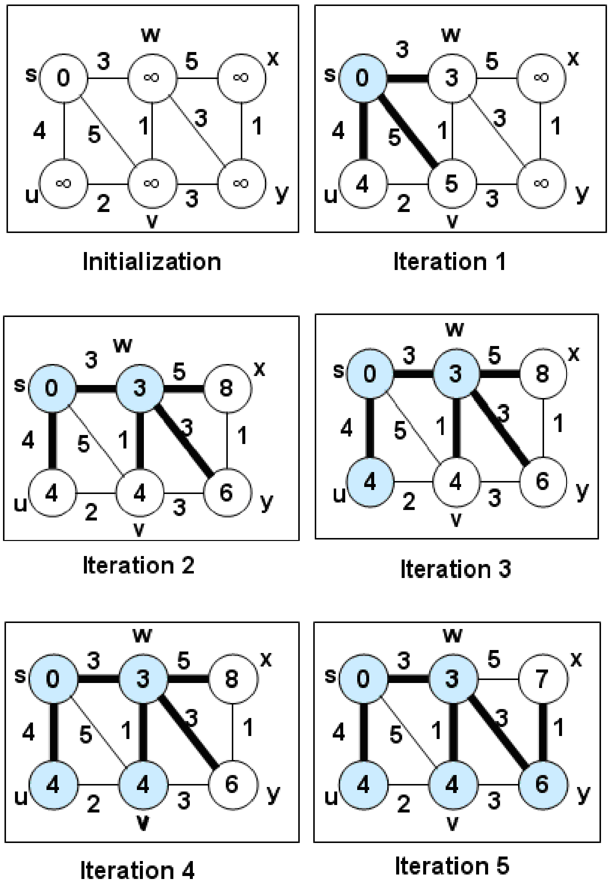

Figure 2: Example to Illustrate the Working of the Dijsktra's Shortest Path Algorithm.

3.2 Smallest Bottleneck Path Problem

In the context of the smallest bottleneck path problem, we define the bottleneck weight of a path *p* to be the maximum of the weights of the constituent edges, $e \in p$. Given the set of all loop-free paths *P* between a source node *s* and destination node *d*, the smallest bottleneck path is the path with the smallest bottleneck weight. We can express this mathematically as $\underset{p\in P}{Min}\left[\underset{\forall e\in p}{Max}\big(weight(e)\big)\right]$.
The

Begin Algorithm *Modified-Dijkstra-Smallest-Bottleneck-Path* (*G*, *s*)
For each vertex $v \in V$
weight [*v*] ← +∞ // an estimate of the smallest bottleneck weight path from *s* to *v*
End For
weight [*s*] ← - ∞
$S \leftarrow \Phi$ // set of nodes for which we know the smallest bottleneck weight path from *s*

$Q \leftarrow V$ // set of nodes for which we know an estimate of the smallest bottleneck weight path from *s*
While $Q \neq \Phi$
  $u \leftarrow$ EXTRACT-MIN (Q)
  $S \leftarrow S \cup \{u\}$
  For each vertex *v* such that $(u, v) \in E$
    If *weight*[*v*] > Max(*weight* [*u*], *weight* (*u*, *v*)) **then**
      weight [*v*] ← Max (*weight* [*u*], *weight* (*u*, *v*))
      Predecessor (*v*) ← *u*
    End If
  End For
End While
End *Modified-Dijkstra-Smallest-Bottleneck-Path*

Figure 3: Pseudo Code for the Modified Dijkstra Smallest Bottleneck Path Algorithm.

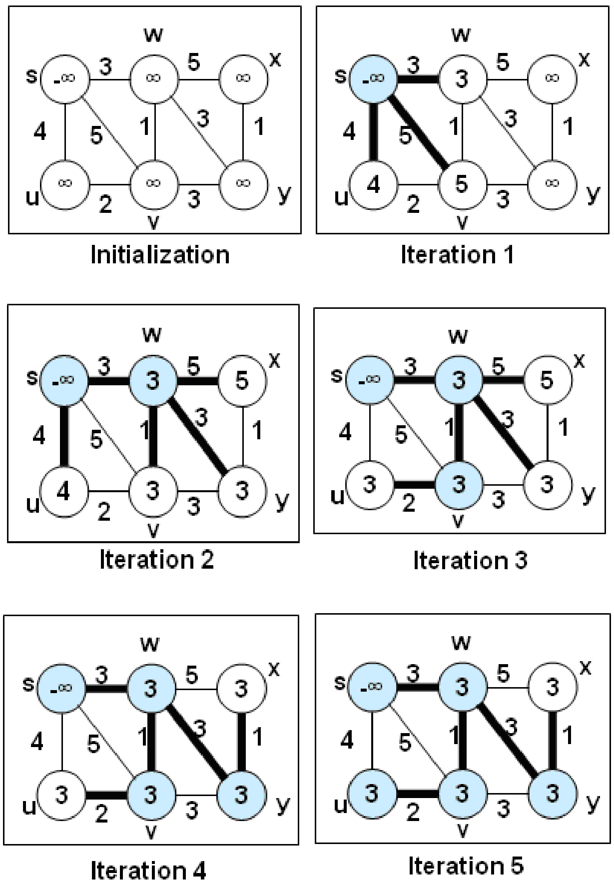

Figure 4: Example for the Smallest Bottleneck Path Problem.

NVSP protocol can be implemented in a graph theoretic context through a modified version of the Dijkstra's algorithm (pseudo code in Figure 3) that solves the smallest bottleneck path problem. Accordingly, the weight of a link from node u to node v is the velocity of the downstream node v. To start with, the weight of the smallest bottleneck path from the source vertex s to every other vertex is estimated to be +∞; whereas the weight of the smallest bottleneck path from the source vertex s to itself is set to - ∞. At the beginning of an iteration, the vertex (say u) with the smallest bottleneck weight among the vertices that have been not yet optimized is now considered to be optimized. As part of the relaxation step, we check whether the current weight of the smallest bottleneck path to any non-optimized neighbor v, i.e. weight[v], is greater than the maximum of the weight of the recently optimized largest bottleneck path from s to u, i.e. weight[u] and the weight of the edge (u, v). If this relaxation condition evaluates to true, then the bottleneck weight of the path from s to v is correspondingly updated (i.e., weight[v] = Max (weight[u], weight(u, v)) and the predecessor of v is set to be u for the path from s to v. This step is repeated over all iterations. A working example is presented in Figure 4. The run-time complexity of the modified Dijkstra algorithm for the smallest bottleneck path problem is the same as that of the original Dijkstra algorithm.

3.3 All Pairs Smallest Bottleneck Paths Problem

In this section, we show that the smallest bottleneck path between any two vertices *u* and $v \in V$ in an undirected weighted graph $G = (V, E)$ is the path between *u* and *v* in the minimum spanning tree of *G*. The Prim's algorithm [4] is a well-known algorithm to determine the minimum spanning tree of weighted graphs and its pseudo code is illustrated in Figure 5. The Prim's algorithm is very similar to the Dijkstra algorithm – the major difference is in the relaxation step.

Begin Algorithm *Prim* (*G*, *s)*; *s* – is any arbitrarily chosen starting vertex
  For each vertex $v \in V$
    weight [*v*] ← ∞
  End For
  weight [*s*] ← 0
  $S \leftarrow \Phi$ // set of nodes whose bottleneck weights will not change further
  $Q \leftarrow V$ // set of nodes whose bottleneck weights are only estimates; final weight could change
  While $Q \neq \Phi$
    $u \leftarrow$ EXTRACT-MIN (Q)
    $S \leftarrow S \cup \{u\}$
    For each vertex *v* such that $(u, v) \in E$
      If *weight* [*v*] > *weight* (*u*, *v*) **then**
      weight [*v*] ← *weight* (*u*, *v*)
      Predecessor (*v*) ← *u*
      End If
    End For
  End While
End *Prim*

Figure 5: Pseudo Code for the Prim's Algorithm to Determine a Minimum Spanning Tree.

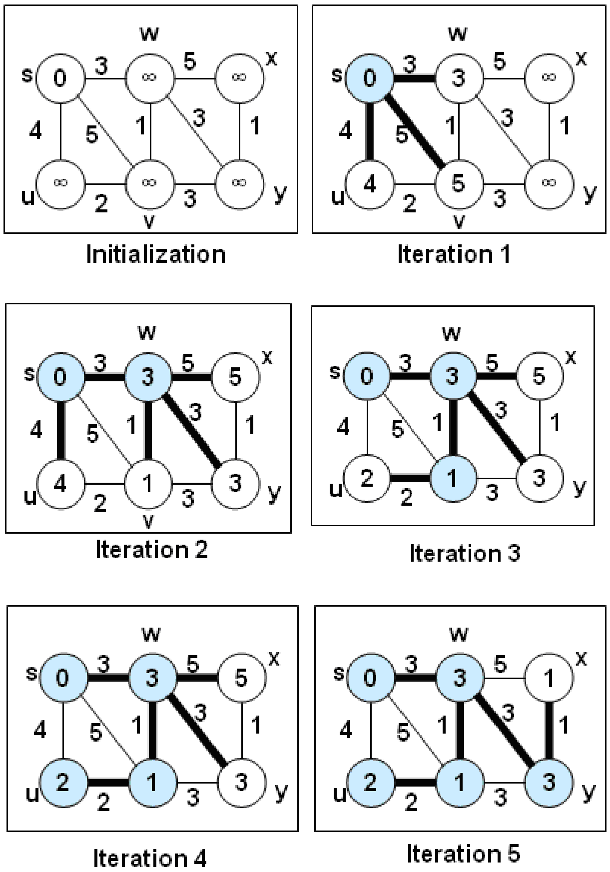

Figure 6: Example for the Minimum Spanning Tree – All Pairs Smallest Bottleneck Paths.

The Prim's algorithm work as follows: The starting vertex is any arbitrarily chosen vertex (say *s*) in the given undirected weighted graph *G*. To begin with, the weights of the smallest bottleneck paths from the starting vertex to every other vertex is assumed to be $+\infty$ (as estimate value, indicating that the paths are actually not known) and the path from the starting vertex to itself is assumed to be 0. During every iteration, we determine the smallest bottleneck path from the starting vertex *s* to a particular vertex *u*, which would be the vertex with the minimum weight among the vertices that have been not yet optimized (i.e. for which the smallest bottleneck path has not been yet determined). We then explore the neighbors of *u* and determine whether we can reach any of the neighbor vertex, say *v*, from *s* through *u* on a path with weight less than the estimated bottleneck weight of the current path we know from *s* to *v*. If we could find such a neighbor *v* as part of the relaxation step, we set the new estimated bottleneck weight of vertex *v* to the weight of the edge (*u*, *v*) and also set the predecessor of *v* to be vertex *u* on the smallest bottleneck path from *s* to *v*. The darkened edges shown in the working example of Figure 6 are the edges that are part of the smallest bottleneck path tree rooted at the starting vertex *s*. The path between any two vertices in this smallest bottleneck path tree is the smallest bottleneck path between the two vertices in the original graph. The run-time complexity of the Prim's minimum spanning tree algorithm is $O(|V|^2)$.

Note that in both Figures 4 and 6, we start with the same initial graph. Since, the relaxation step of the modified Dijkstra algorithm and the Prim's algorithm are different, the sequence of vertices that are optimized in each algorithm is different from one another. However, the final tree rooted at the starting vertex *s* is the same in both the figures. This example vindicates our argument that the minimum spanning tree contains the smallest bottleneck paths between any two vertices in the original graph. We now formally prove this argument (refer Figure 7 for an illustration of the example) below through the method of Proof by Contradiction.

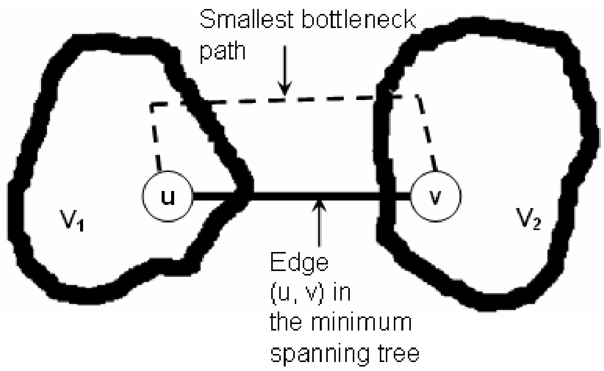

Figure 7: Proof by Contradiction: Minimum Spanning Tree with All Pairs Smallest Bottleneck Paths.

Let there be a pair of vertices $u \in V_1$ and $v \in V_2$ in $G = (V, E)$ such that the edge $(u, v) \in E$, $V_1 \cup V_2 = V$ and $V_1 \cap V_2 = \Phi$. Assume the edge (*u*, *v*) belongs to the minimum spanning tree *T* of *G*; but the edge is not part of the smallest bottleneck path from *u* to *v*. Let there exist an alternate path from *u* to *v* that is the smallest bottleneck path. Since *u* and *v* are in two disjoint vertex partitions, there should be at least one edge (call it *e*) in the path from *u* to *v* with endpoint vertices in each partition. But, by definition of a minimum spanning tree, the weight(*u*, *v*) $\leq$ weight(edge *e*); otherwise, a cheaper tree could be obtained by replacing (*u*, *v*) with the edge *e* and *T* containing (*u*, *v*) would not be a minimum spanning tree. Hence, edge *e* could be replaced by edge (*u*, *v*) without increasing the weight of any smallest bottleneck path. Likewise, we can prove that every edge in *T* would be part of a smallest bottleneck path. Since *T* is a minimum spanning tree, all its edges constitute the all pairs smallest bottleneck paths for the entire graph.

3.4 Largest Bottleneck Paths Problem

In the context of the largest bottleneck path problem, we define the bottleneck weight of a path *p* to be the minimum of the weights of the constituent edges, $e \in p$. Given the set of all loop-free paths *P* between a source node *s* and destination node *d*, the largest bottleneck path is the path with the largest bottleneck weight. We can express this mathematically as $\underset{p \in P}{Max}\left[\underset{\forall e \in p}{Min}\left(weight(e)\right)\right]$.

```
Begin Algorithm Modified-Dijkstra-Largest-Bottleneck-
                    Path (G, s)
   For each vertex v Є V
       weight [v] ← - ∞ // an estimate of the largest
                        bottleneck weight path from s to v
   End For
   weight [s] ← + ∞
   S ← Φ // set of nodes for which we know the largest
            bottleneck weight path from s
   Q ← V // set of nodes for which we know an
            estimate of the largest bottleneck weight
            path from s
   While Q ≠ Φ
     u ← EXTRACT-MAX (Q)
     S ← S U {u}
     For each vertex v such that (u, v) Є E
       If weight[v] < Min (weight [u], weight (u, v)) then
          weight [v] ← Min (weight [u], weight (u, v))
          Predecessor (v) ← u
        End If
     End For
   End While
End Modified-Dijkstra-Largest-Bottleneck-Path
```

Figure 8: Pseudo Code for the Modified Dijkstra's Algorithm for the Largest Bottleneck Path Problem.

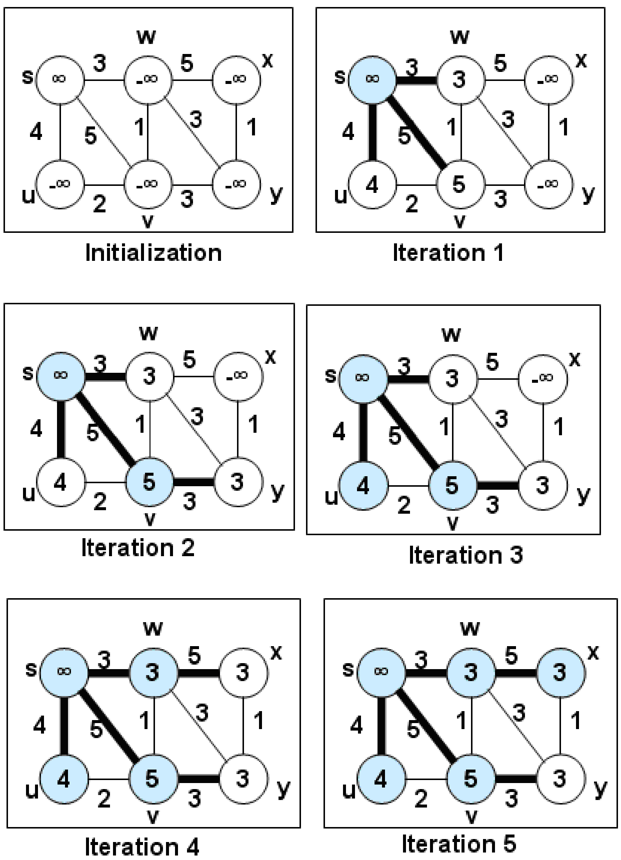

Figure 9: Example for the Largest Bottleneck Path Problem.

The FORP protocol can be simulated using a modified version of the Dijsktra's algorithm (pseudo code in Figure 8) that solves the Largest Bottleneck Path problem on a static graph. The edge weights correspond to the predicted LET values for the corresponding links. To start with, the weight of the largest bottleneck path from the source vertex *s* to every other vertex is estimated to be - ∞; whereas the weight of the largest bottleneck path from the source vertex *s* to itself is set to +∞. At the beginning of an iteration, the vertex (say *u*) with the largest bottleneck weight among the vertices that have been not yet optimized is now considered to be optimized (i.e., the largest bottleneck path from the source vertex *s* to the vertex *u* is considered to have been determined by now). As part of the relaxation step, we check whether the current weight of the largest bottleneck path to any non-optimized neighbor *v*, i.e. *weight*[*v*], is lower than the minimum of the weight of the recently optimized largest bottleneck path from *s* to *u*, i.e. *weight*[*u*] and the weight of the edge (*u*, *v*). If this relaxation condition evaluates to true, then the bottleneck weight of the path from *s* to *v* is correspondingly updated (i.e., *weight*[*v*] = Min(*weight*[*u*], *weight*(*u*, *v*)) and the predecessor of *v* is set to be *u* for the path from *s* to *v*. This step is repeated over all iterations. A working example is presented in Figure 9. The run-time complexity of the modified Dijkstra algorithm for the largest bottleneck path problem is the same as that of the original algorithm for the shortest path problem.

3.5 All Pairs Largest Bottleneck Paths Problem

In this section, we show that the largest bottleneck path between any two vertices *u* and $v \in V$ in an undirected weighted graph $G = (V, E)$ is the path between *u* and *v* in the maximum spanning tree of *G*. The maximum spanning tree of a graph can be determined using a slightly modified version of the Prim's algorithm – the modification is in the initialization step and the relaxation condition. In the original Prim's algorithm, the initial weight of all the vertices other than the starting vertex is set to +∞; whereas in the modified Prim's algorithm for the all pairs largest bottleneck path problem, the initial weight of all the vertices other than the starting vertex is set to – ∞ (an initial estimate for the largest bottleneck paths, which are actually not know to start with). The weight of the starting vertex, *s*, in both algorithms is 0. The pseudo code of the modified Prim's algorithm is given in Figure 10.

```
Begin Algorithm Modified-Prim (G, s); s – is any
                  arbitrarily chosen starting vertex
   For each vertex v Є V
       weight [v] ← – ∞
   End For
   weight [s] ← 0
   S ← Φ // set of nodes whose bottleneck weights
               will not change further
   Q ← V // set of nodes whose bottleneck weights
```

```
                are only estimates and the final weight
                could change
While Q ≠ Φ
   u ← EXTRACT-MAX (Q)
   S ← S U {u}
    For each vertex v such that (u, v) Є E
       If weight [v] < weight (u, v) then
       weight [v] ← weight (u, v)
       Predecessor (v) ← u
       End If
    End For
 End While
End Modified-Prim
```

Figure 10: Pseudo Code for the Modified Prim's Algorithm to Determine a Maximum Spanning Tree

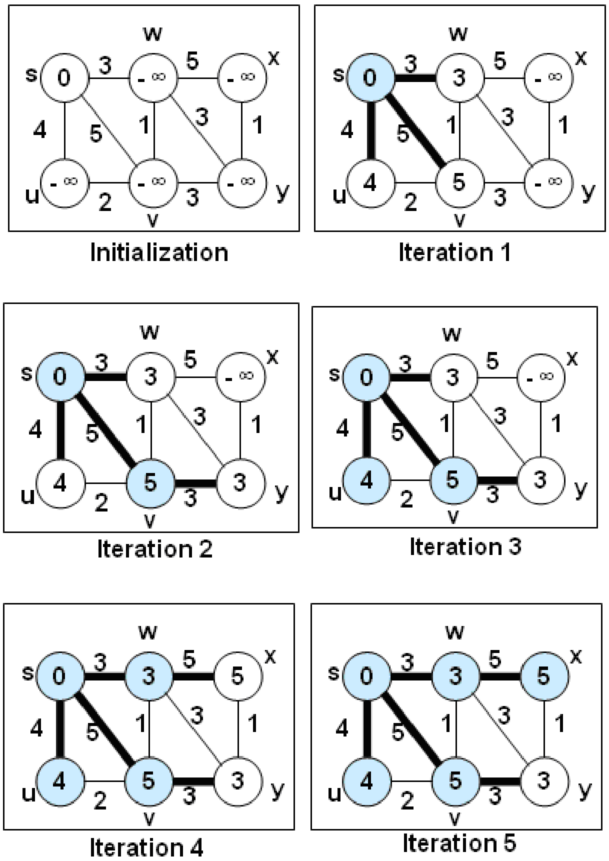

Figure 11: Example for the Maximum Spanning Tree – All Pairs Largest Bottleneck Paths.

During every iteration, we determine the largest bottleneck path from the starting vertex *s* to a particular vertex *u*, which would be the vertex with the maximum weight among the vertices that have been not yet optimized (i.e. for which the largest bottleneck path has not been yet determined). We then explore the neighbors of *u* and determine whether we can reach any of the neighbor vertices, say *v*, from *s* through *u* on a path with weight greater than the estimated bottleneck weight of the current path we know from *s* to *v*. If we could find such a neighbor *v* as part of the relaxation step, we set the new estimated bottleneck weight of vertex *v* to the weight of the edge (*u*, *v*) and also set the predecessor of *v* to be vertex *u* on the largest bottleneck path from *s* to *v*. The darkened edges shown in the working example of Figure 11 are the edges that are part of the largest bottleneck path tree rooted at the starting vertex *s*. The path between any two vertices in this largest bottleneck path tree is the largest bottleneck path between the two vertices in the original graph. The run-time complexity of the modified Prim's algorithm to determine maximum spanning tree is $O(|V|^2)$, the same as the original Prim's algorithm for minimum spanning trees. The correctness of the modified Prim's algorithm for the all pairs largest bottleneck path problem can be proved using the same logic used to prove the correctness of the Prim's algorithm for the all pairs smallest bottleneck path problem.

4 Graph Theory Algorithms for Multicast Communication in MANETs

The original Dijkstra shortest path algorithm described previously can be used to determine shortest path trees with minimum hop count per source-receiver path. The problem of determining a multicast tree with the minimum number of links is a NP-complete problem for which there is no polynomial-time algorithm yielding an optimistic solution. Hence, algorithms have been proposed in the literature to approximate such multicast trees. The Steiner tree algorithm is the best-known algorithm in the literature to approximate multicast trees with the minimum number of links connecting a source node to the receiver nodes of the multicast group.

Given a static graph, $G = (V, E)$, where V is the set of vertices, E is the set of edges and a subset of vertices (called the multicast group or Steiner points) $MG \subseteq V$, the multicast Steiner tree is the tree with the least number of edges required to connect all the vertices in MG. In this paper, we use a well-known $O(|V||MG|^2)$ algorithm of Kou et al [9], where $|V|$ is the number of nodes in the network graph and $|MG|$ is the size of the multicast group comprising of the source nodes and the receiver nodes, to approximate the minimum edge Steiner tree in graphs representing snapshots of the network topology. In unit disk graphs such as the static graphs used in our research, Step 5 of the algorithm (pseudo code in Figure 12) is not needed and the minimal spanning tree T_{MG} obtained at the end of Step 4 could be considered as the minimum edge Steiner tree. One could use the Prim's algorithm to find the minimum spanning trees.

Input: A Static Graph $G = (V, E)$
Multicast Group $MG \subseteq V$
Output: A *MG-Steiner-tree* for the set $MG \subseteq V$

Begin Kou et al Algorithm (G, MG)

Step 1: Construct a complete undirected weighted graph $G_C = (MG, E_C)$ from G and MG where $\forall (v_i, v_j) \in E_C$, v_i and v_j are in MG, and the weight of edge (v_i, v_j) is the length of the shortest path from v_i to v_j in G.
Step 2: Find the minimum weight spanning tree T_C in G_C (If more than one minimal spanning tree exists, pick an arbitrary one).
Step 3: Construct the sub graph G_{MG} of G, by replacing each edge in T_C with the corresponding shortest path from G (In case of any tie, between an arbitrary shortest path between the two vertices).
Step 4: Find the minimal spanning tree T_{MG} in G_{MG} with unit edge weights. If more than one minimal spanning tree exists, pick an arbitrary one).
return T_{MG} as the *MG-Steiner-tree*

End *Kou et al Algorithm*

Figure 12: Kou et al's Algorithm to Find an Approximate Minimum Edge Steiner Tree.

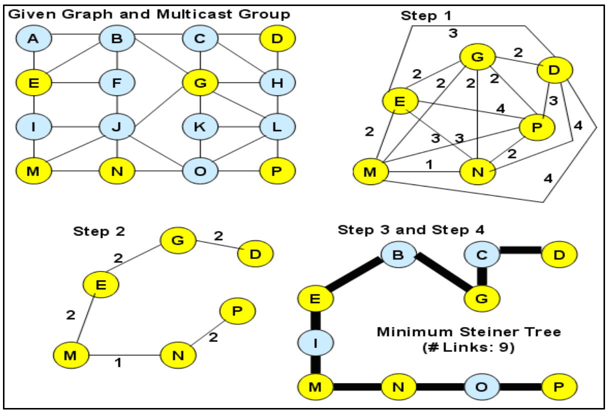

Figure 13: Construction of a Minimum Steiner Tree using Kou et al.'s Algorithm.

We now illustrate the working of the Kou et al's algorithm through an example in Figure 13. The vertices {D, G, E, M, N, P} form the multicast group in the vertex set {A, B … P}. As observed in the example, the sub graph G_{MG} obtained in Step 3 is nothing but the minimal spanning tree T_{MG}, which is the output of Step 4. In general, for unit disk graphs, like the static graphs we are working with, the outputs of both Steps 3 and 4 are the same and it is enough that we stop at Step 3 and output the MG-Steiner-tree.

5 Graph Theory Algorithms for Broadcast Communication in MANETs

In this section, we describe a maximum-density based CDS (MaxD-CDS) graph theoretic algorithm to approximate a Minimum Connected Dominating Set (MCDS) in a static graph, taken as a snapshot of a MANET topology.

5.1 Data Structures used by the MaxD-CDS Algorithm and Breadth First Search

We use the following principal data structures for the MaxD-CDS algorithm:

(i) *CDS-Node-List* – includes all nodes that are members of the CDS
(ii) *Covered-Nodes-List* – includes all nodes that are in the CDS-Node-List and all nodes that are adjacent to at least one member of the CDS-Node-List.

Before we run the CDS formation algorithm, we make sure the underlying network graph is connected by running the Breadth First Search (BFS) algorithm [4]; because, if the underlying network graph is not connected, we would not be able to find a CDS that will cover all the nodes in the network. We run BFS, starting with an arbitrarily chosen node in the network graph. If we are able to visit all the vertices in the graph, then the corresponding network is said to be connected. If the graph is not connected, we simply continue with the static graph (snapshot of the network topology) collected at the next time instant and start with the BFS test. The pseudo code for BFS is shown in Figure 14. The run-time complexity of BFS is $O(|V|+|E|)$.

Input: Graph $G = (V, E)$
Auxiliary Variables/Initialization:
Nodes-Explored = Φ, *FIFO-Queue* = Φ
Begin Algorithm *BFS* (*G*, *s*)
root-node = randomly chosen vertex in *V*
Nodes-Explored = *Nodes-Explored* U {*root-node*}
FIFO-Queue = *FIFO-Queue* U {*root-node*}
while (|*FIFO-Queue*| > 0) **do**
front-node u = Dequeue(*FIFO-Queue*) // extract the first node
for (every edge (*u*, *v*)) **do** // i.e. every neighbor *v* of node *u*
if ($v \notin$ *Nodes-Explored*) **then**
Nodes-Explored = *Nodes-Explored* U {*v*}
FIFO-Queue = *FIFO-Queue* U {*v*}
Parent (*v*) = *u*
end if
end for
end while
if (| *Nodes-Explored* | = | *V* |) **then**
return Connected Graph - true
else return Connected Graph - false
end if
End Algorithm *BFS*

Figure 14: Pseudo Code for the BFS Algorithm to Determine Network Connectivity.

5.2 Maximum Density-based Algorithm to Approximate a MCDS

The idea of the maximum density (MaxD)-based CDS formation algorithm is to select nodes with larger number of uncovered neighbors for inclusion in the CDS. The algorithm forms and outputs a CDS based on a given input graph representing a snapshot of the MANET at a particular time instant. Specifically, the algorithm outputs a list (*CDS-Node-List)* of all nodes that are part of the CDS formed based on the given MANET. The first node to be included in the *CDS-Node-List* is the node with the maximum number of uncovered neighbors (any ties are broken arbitrarily). A CDS member is considered to be "covered", so a CDS member is additionally added to the *Covered-Nodes-List* when it is included in the *CDS-Node-List*. All nodes that are adjacent to a CDS member are also said to be covered, so the uncovered neighbors of a CDS member are also added to the *Covered-Nodes-List* as the member is added to the *CDS-Node-List*. To determine the next node to be added to the *CDS-Node-List*, we must select the node with the largest density amongst the nodes that meet the criteria for inclusion into the CDS.

Input: Graph $G = (V, E)$, where V is the vertex set and E is the edge set.
Source vertex, s – vertex with the largest number of uncovered neighbors in V.
Auxiliary Variables and Functions: *CDS-Node-List*, *Covered-Nodes-List, Neighbors*(v), $\forall\, v \in V$.
Output: *CDS-Node-List*
Initialization: *Covered-Nodes-List* = $\{s\}$,
CDS-Node-List = Φ
Begin Construction of *MaxD-CDS*
while (|*Covered-Nodes-List*| < $|V|$) **do**
Select a vertex $r \in$ *Covered-Nodes-List* and $r \notin$ *CDS-Node-List* such that r has the largest number of uncovered neighbors that are not in *Covered-Nodes-List*

CDS-Node-List = *CDS-Node-List* U $\{r\}$

for all $u \in$ *Neighbors*(r) and $u \notin$ *Covered-Nodes-List*
Covered-Nodes-List = *Covered-Nodes-List* U $\{u\}$
end for
end while
return *CDS-Node-List*
End Construction of *MaxD-CDS*

Figure 15: Pseudo Code for the Algorithm to Construct the Maximum Density (MaxD)-based CDS.

The criteria for CDS membership selection are the following: the node should not already be a part of the CDS (*CDS-Node-List),* the node must be in the *Covered-Nodes-List*, and the node must have at least one uncovered neighbor (at least one neighbor that is not in the *Covered-Nodes-List*). Amongst the nodes that meet these criteria for CDS membership inclusion, we select the node with the largest density (i.e., the largest number of uncovered neighbors) to be the next member of the CDS. Ties are broken arbitrarily. This process is repeated until all nodes in the network are included in the *Covered-Nodes-List*. Once all nodes in the network are considered to be "covered", the CDS is formed and the algorithm returns a list of nodes in the resulting MaxD-CDS (nodes in the *CDS-Node-List*). The run-time complexity of the MaxD-CDS algorithm is $O(|V|^2+|E|)$ since there would be at most $|V|$ iterations – it would take $O(|V|)$ time to determine the node with the largest number of uncovered neighbors in each iteration and across all these iterations, we would be visiting $|E|$ edges totally. The pseudo code for the MaxD-CDS algorithm is given in Figure 15 and a working example of the algorithm is illustrated in Figure 16.

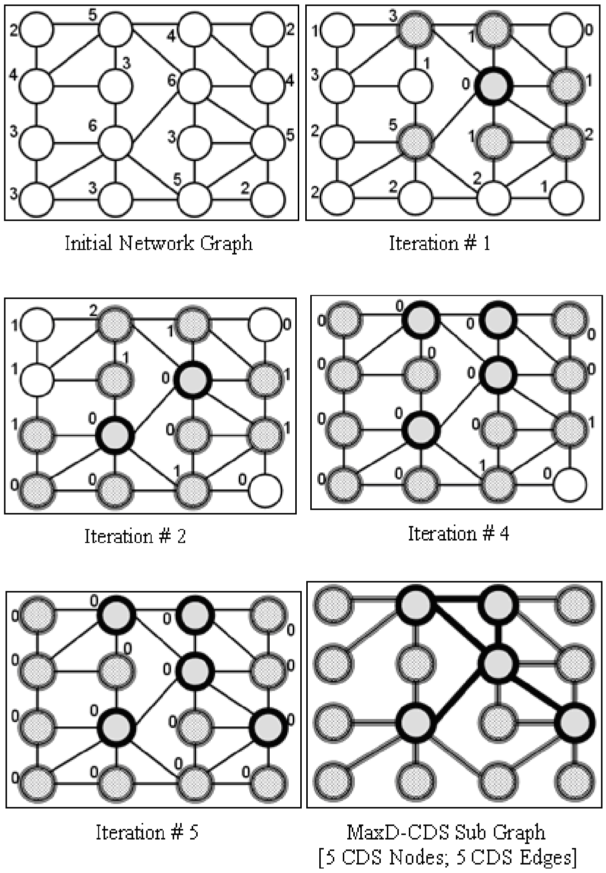

Figure 16: Example to Illustrate the Construction of a Maximum Density (MaxD)-based CDS

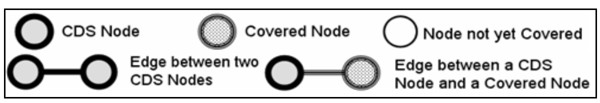

Figure 17: Legend for Figure 16

6 Graph Theory Algorithms for Multi-path Communication in MANETs

Let P_L, P_N and P_Z be the set of link-disjoint, node-disjoint and zone-disjoint *s*-*d* routes respectively. We use the Dijkstra O($|V|^2$) algorithm to determine the minimum hop *s*-*d* path in a graph of *n* nodes. We assume the *s*-*d* routes in a multi-path set are used in the increasing order of the hop count. In other words, the *s*-*d* route with the least hop count is used as long as it exists, then the *s*-*d* route with the next highest hop count is used as long as it exists and so on. We thus persist with the determined multi-path set of *s*-*d* routes as long as at least one path in the set exists.

6.1 Algorithm to Determine Link-Disjoint Paths

To determine the set of link-disjoint paths, P_L, (refer Figure 18), we remove all the links that were part of *p* from the graph *G* to obtain a modified graph G^L (V, E^L). We then determine the minimum hop *s*-*d* path in the modified graph *G*', add it to the set P_L and remove the links that were part of this path to get a new updated G^L (V, E^L). We repeat this procedure until there exists no more *s*-*d* paths in the network. The set P_L is now said to have the link-disjoint *s*-*d* paths in the original network graph *G* at the given time instant. Figure 20 illustrates a working-example of the algorithm to find the set of link-disjoint paths on a static graph.

Input: Graph G (V, E), source vertex *s* and destination vertex *d*
Output: Set of link-disjoint paths P_L
Auxiliary Variables: Graph G^L (V, E^L)
Initialization: $G^L (V, E^L) \leftarrow G (V, E)$, $P_L \leftarrow \Phi$
Begin *Algorithm Link-Disjoint-Paths*

while ($\exists$ at least one *s*-*d* path in G^L)
  $p \leftarrow$ Minimum hop *s*-*d* path in G^L.
  $P_L \leftarrow P_L \cup \{p\}$
  $\forall_{edge, e \in p}\ G^L (V, E^L) \leftarrow G^L (V, E^L - \{e\})$
end While
return P_L
End *Algorithm Link-Disjoint-Paths*

Figure 18: Algorithm to Determine the Set of Link-Disjoint *s*-*d* Paths in a Network Graph.

6.2 Algorithm to Determine Node-Disjoint Paths

To determine the set of node-disjoint paths, P_N, (refer Figure 19), we remove all the intermediate nodes (nodes other than the source vertex *s* and destination vertex *d*) that were part of the minimum hop *s*-*d* path *p* in the original graph *G* to obtain the modified graph, G^N (V^N, E^N). We determine the minimum hop *s*-*d* path in the modified graph G^N (V^N, E^N), add it to the set P_N and remove the intermediate nodes that were part of this *s*-*d* path to get a new updated G^N (V^N, E^N). We then repeat this procedure until there exists no more *s*-*d* paths in the network. The set P_N is now said to contain the node-disjoint *s*-*d* paths in the original network graph *G*. Figure 21 illustrates a working example of the algorithm to find the set of node-disjoint paths on a static graph.

Input: Graph G (V, E), source vertex *s* and destination vertex *d*
Output: Set of node-disjoint paths P_N
Auxiliary Variables: Graph G^N (V^N, E^N)
Initialization: $G^N (V^N, E^N) \leftarrow G (V, E)$, $P_N \leftarrow \Phi$
Begin *Algorithm Node-Disjoint-Paths*

While ($\exists$ at least one *s*-*d* path in G^N)
  $p \leftarrow$ Minimum hop *s*-*d* path in G^N.
  $P_N \leftarrow P_N \cup \{p\}$
  $\forall_{\substack{vertex, v \in p \\ v \neq s,d \\ edge, e \in Adj-list(v)}}\ G^N (V^N, E^N) \leftarrow G^N (V^N - \{v\}, E^N - \{e\})$
end While
return P_N
End *Algorithm Node-Disjoint-Paths*

Figure 19: Algorithm to Determine the Set of Node-Disjoint *s*-*d* Paths in a Network Graph.

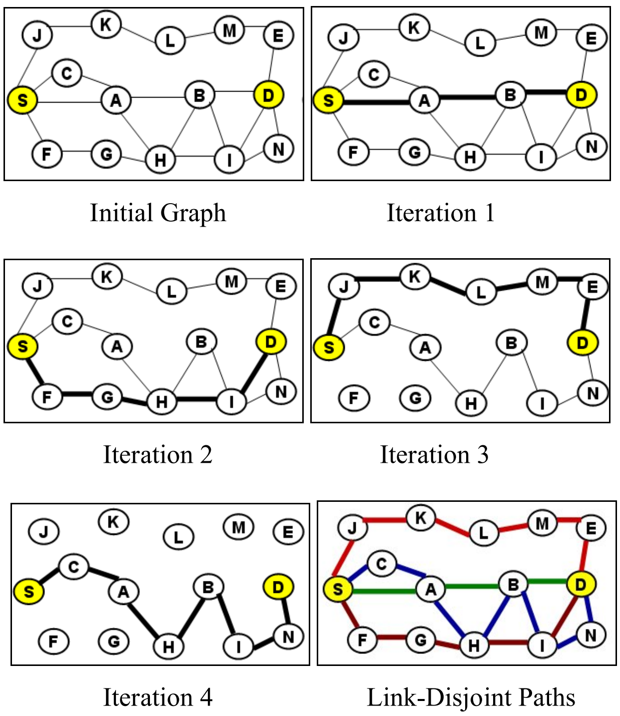

Figure 20: Example to Illustrate the Working of the Algorithm to Find Link-Disjoint Paths.

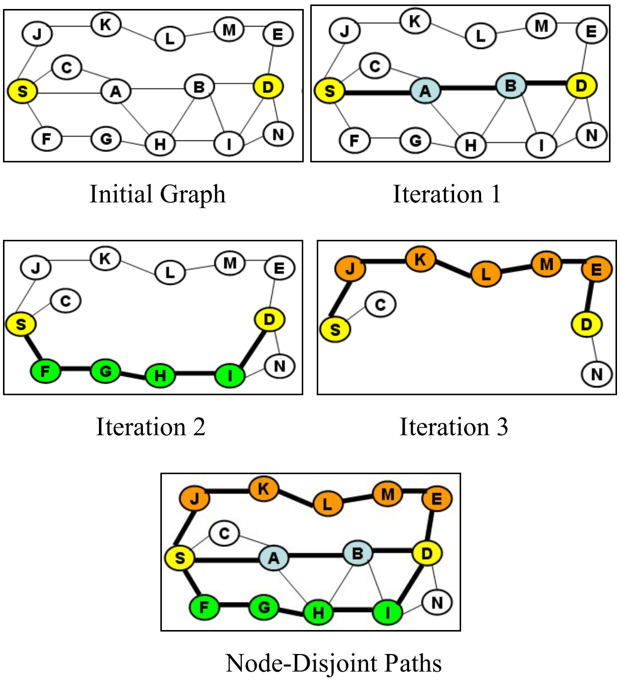

Figure 21: Example to Illustrate the Working of the Algorithm to Find Node-Disjoint Paths.

6.3 Algorithm to Determine Zone-Disjoint Paths

To determine the set of zone-disjoint paths, P_Z, (refer Figure 22), we remove all the intermediate nodes (nodes other than the source vertex *s* and destination vertex *d*) that were part of the minimum hop *s-d* path *p* and also all their neighbor nodes from the original graph *G* to obtain the modified graph G^Z (V^Z, E^Z). We determine the minimum hop *s-d* path in the modified graph G^Z, add it to the set P_Z and remove the intermediate nodes that were part of this *s-d* path and all their neighbor nodes to obtain a new updated graph G^Z (V^Z, E^Z). We then repeat this procedure until there exists no more *s-d* paths in the network. The set P_Z is now said to contain the set of zone-disjoint *s-d* paths in the original network graph *G*. Note that when we remove a node *v* from a network graph, we also remove all the links associated with the node (i.e., links belonging to the adjacency list *Adj-list*(*v*)) whereas when we remove a link from a graph, no change occurs in the vertex set of the graph. Figure 23 illustrates a working example of the algorithm to find the set of zone-disjoint paths on a static graph.

Input: Graph *G* (*V*, *E*), Source vertex *s* and Destination vertex *d*
Output: Set of Zone-Disjoint Paths P_Z
Auxiliary Variables: Graph G^Z (V^Z, E^Z)
Initialization: $G^Z (V^Z, E^Z) \leftarrow G (V, E)$, $P_Z \leftarrow \Phi$
Begin *Algorithm Zone-Disjoint-Paths*
1 **While** ($\exists$ at least one *s-d* path in G^Z)
2 $p \leftarrow$ Minimum hop *s-d* path in G^Z
3 $P_Z \leftarrow P_Z \cup \{p\}$
4 $\forall_{\substack{vertex, u \in p, u \neq s,d \\ edge, e \in Adj-list(u)}}$ $G^Z (V^Z, E^Z) \leftarrow G^Z (V^Z - \{u\}, E^Z - \{e\})$
5 $\forall_{\substack{vertex, u \in p, u \neq s,d \\ v \in Neighbor(u), v \neq s,d \\ edge, e' \in Adj-list(v)}}$ $G^Z (V^Z, E^Z) \leftarrow G^Z(V^Z - \{v\}, E^Z - \{e'\})$
6 **end While**
7 **return** P_Z
End *Algorithm Zone-Disjoint-Paths*

Figure 22: Algorithm to Determine the Set of Zone-Disjoint *s-d* Paths in a Network Graph.

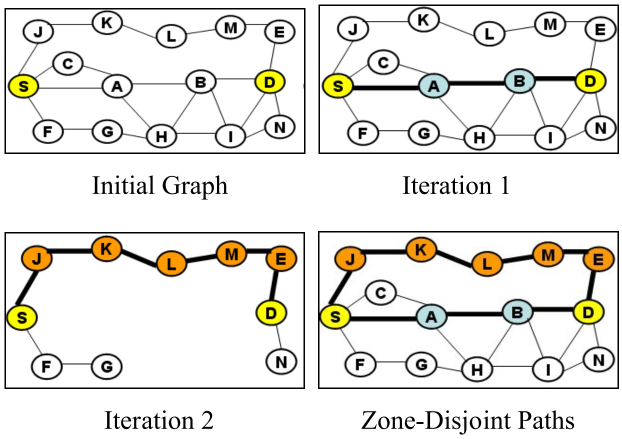

Figure 23: Example to Illustrate the Working of the Algorithm to Find Zone-Disjoint Paths

7 Conclusions

The high-level contribution of this paper is the idea of using traditional graph theory algorithms, which have been taught in academic institutions at undergraduate and graduate level, to simulate and study the behavior of the complex routing protocols for unicast, multicast, broadcast and multi-path communication in MANETs. In the Section on Background work, we provided an exhaustive set of background information on the routing protocols that have been proposed for the above different communication problems. In the subsequent sections, we described one or more graph theoretic algorithms for studying each of these communication problems. We chose the Dijkstra algorithm for shortest path routing as the core algorithm and meticulously modified it and/or adopted it to (i) find a solution for the largest bottleneck path and smallest bottleneck path problems, which could be used to determine a sequence of stable routes as well as to (ii) find a set of link-disjoint, node-disjoint or zone-disjoint routes for multi-path communication. We illustrate the use of Prim's algorithm for minimum spanning tree to determine the 'All Pairs Smallest Bottleneck Paths' and also show how the modified version of the Prim's algorithm to determine maximum spanning tree can be used to solve the 'All Pairs Largest Bottleneck Paths' problem. We prove that the path between any two nodes in the minimum spanning tree is the smallest bottleneck path between those two nodes in the original graph. We also discussed a maximum density-based algorithm to approximate minimum connected dominating sets (MCDS) for MANETs such that the MCDS could be used as a backbone towards broadcast communication for route discoveries and other global communication needs. In addition, we discussed the Steiner tree problem and the Kou et al.'s algorithm to approximate a multicast tree that has the minimum number of links connecting the source nodes to the members of a multicast group. Each of the graph theoretic algorithms discussed in this paper presents the optimal solutions or the best approximations for the appropriate problems they have been suggested for and all of them could be implemented in the most efficient manner using the pseudo code presented and executed in a polynomial run-time.

As a concluding remark, we would like to state that the proposed idea could go a long way in avoiding the significant loss of time faced by student researchers to understand and modify the simulation code for even conducting simple experimental studies. This idea could be adopted to facilitate undergraduate student research in the area of MANETs without requiring the students to directly work on the complex discrete-event simulators. A successful implementation of this idea is the Research Experiences for Undergraduates (REU) site in the areas of Wireless Ad hoc Networks and Sensor Networks, hosted by the Department of Computer Science at Jackson State University, Jackson, MS, USA. The REU site is currently being funded by the U.S. National Science Foundation (NSF) and is accessible through http://www.jsums.edu/cms/reu.

8 Future Research Directions

Graph theory algorithms form the backbone for research on communication protocols for wireless ad hoc networks and sensor networks. This paper lays the foundation for use of several simplistic graph theoretic algorithms (taught at the undergraduate and graduate level) to simulate the behavior of the complex MANET routing protocols. The next step of research in this direction would involve implementing these graph theoretic algorithms in a centralized environment using offline traces of the mobility profiles of the nodes (under a particular mobility model) to generate the mobile graph (i.e., sequence of static graphs representing snapshots of the network topology at different time instants) and compare the performance metrics obtained for the communication structures with that of those obtained for the actual routing protocols when simulated in a discrete-event simulator such as ns-2, GloMoSim and etc. Some of the performance metrics that could be directly compared are the hop count per source-destination path (for unicasting), hop count per source-receiver path (for multicasting), number of links per multicast tree, lifetime per path, lifetime per multicast tree, time between two consecutive route discoveries for link-disjoint, node-disjoint and zone-disjoint routes, number of nodes per CDS, hop count of a source-destination path per CDS and etc. We conjecture that the results obtained for the above performance metrics from the centralized graph theory implementations will serve as the optimal benchmarks to which the results obtained from the actual routing protocols in a discrete-event simulator environment would be actually bounded under. This is because the centralized implementations would assume an ideal medium-access control (MAC) layer that would not offer any interference to constrain the communication.

If the simulations could be conducted in more than one discrete-event simulator, then the results for the performance metrics obtained from the different simulators could be compared to the optimal benchmarks obtained with our theoretical algorithms and could be helpful in identifying the simulator that gives performance closest to the optimum for a particular communication problem (unicast, multicast, broadcast, multi-path) under specific operating conditions. Our proposed approach of using graph theory algorithms to study the MANET routing protocols could also be extended to wireless sensor networks, wherein we can use the tree and CDS construction algorithms to study the data gathering protocols.

Acknowledgments

Research was sponsored by the U. S. National Science Foundation grant (CNS-0851646) entitled: "REU Site: Undergraduate Research Program in Wireless Ad hoc Networks and Sensor Networks. The material for this journal paper evolved from the tutorial slides developed by the author to simulate the routing protocols for mobile ad hoc networks with centralized and distributed implementations of the appropriate graph theory algorithms. The views and conclusions in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the funding agency.

References

[1] K. M. Alzoubi, P.-J Wan and O. Frieder, "Distributed Heuristics for Connected Dominating Set in Wireless Ad Hoc Networks," *IEEE / KICS Journal on Communication Networks*, Vol. 4, No. 1, pp. 22-29, 2002.

[2] S. Butenko, X. Cheng, D.-Z. Du and P. M. Paradlos, "On the Construction of Virtual Backbone for Ad Hoc Wireless Networks," *Cooperative Control: Models, Applications and Algorithms*, pp. 43-54, Kluwer Academic Publishers, 2002.

[3] S. Butenko, X. Cheng, C. Oliviera and P. M. Paradlos, "A New Heuristic for the Minimum Connected Dominating Set Problem on Ad Hoc Wireless Networks," *Recent Developments in Cooperative Control and Optimization*, pp. 61-73, Kluwer Academic Publishers, 2004.

[4] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, "Introduction to Algorithms," 2nd Edition, MIT Press/ McGraw-Hill, September 2001.
[5] K. Fall and K. Varadhan, "ns notes and documentation," The VINT Project at LBL, Xerox PARC, UCB, and USC/ISI, http://www.isi.edu/nsnam/ns, August 2001.
[6] B. Hofmann-Wellenhof, H. Lichtenegger and J. Collins, *Global Positioning System: Theory and Practice*, 5th ed., Springer, September 2004.
[7] N. T. Javan and M. Dehghan, "Reducing End-to-End Delay in Multi-path Routing Algorithms for Mobile Ad hoc Networks," *Proceedings of the International Conference on Mobile Ad hoc and Sensor Networks* (MSN 2007), Lecture Notes in Computer Science (LNCS) 4864, pp. 703 – 712, December 2007.
[8] D. B. Johnson, D. A. Maltz and J. Broch, "DSR: The Dynamic Source Routing Protocol for Multihop Wireless Ad Hoc Networks," *Ad Hoc Networking*, edited by Charles E. Perkins, Chapter 5, pp. 139 – 172, Addison Wesley, 2001.
[9] L. Kou, G. Markowsky and L. Berman, "A Fast Algorithm for Steiner Trees," *Acta Informatica*, vol. 15, no. 2, pp. 141-145, 1981.
[10] F. Kuhn, T. Moscibroda and R. Wattenhofer, "Unit Disk Graph Approximation," *Proceedings of the Joint Workshop on Foundations of Mobile Computing*, pp. 17-23, October 2004.
[11] S. Lee and M. Gerla, "Split Multipath Routing with Maximally Disjoint Paths in Ad Hoc Networks," *Proceedings of the IEEE International Conference on Communications*, Vol. 10, pp. 3201-3205, 2001.
[12] N. Meghanathan, "A Beaconless Node Velocity-based Stable Path Routing Protocol for Mobile Ad hoc Networks," *Proceedings of the IEEE Sarnoff Symposium Conference*, Princeton, NJ, March 30-April 1, 2009.
[13] N. Meghanathan, "Exploring the Stability-Energy Consumption-Delay-Network Lifetime Tradeoff of Mobile Ad Hoc Network Routing Protocols," *Journal of Networks*, vol. 3, no. 2, pp. 17 – 28, February 2008.
[14] N. Meghanathan, "Benchmarks and Tradeoffs for Minimum Hop, Minimum Edge and Maximum Lifetime per Multicast Tree in Mobile Ad hoc Networks," *International Journal of Advancements in Technology*, vol. 1, no. 2, pp. 234-251, October 2010.
[15] E. Royer and C. E. Perkins, "Multicast Operation of the Ad-hoc On-demand Distance Vector Routing Protocol," *Proceedings of the 5th ACM/IEEE Annual Conference on Mobile Computing and Networking*, pp. 207-218, Seattle, USA, August 1999.
[16] C. E. Perkins and E. M. Royer, "Ad Hoc On-demand Distance Vector Routing," *Proceedings of the 2nd Annual IEEE International Workshop on Mobile Computing Systems and Applications*, pp. 90 – 100, February 1999.
[17] T. Ozaki, J-B. Kim and T. Suda, "Bandwidth-Efficient Multicast Routing for Multihop, Ad hoc Wireless Networks," *Proceedings of the IEEE INFOCOM Conference*, vol. 2, pp. 1182-1192, Anchorage, USA, April 2001.
[18] W. Su, S-J. Lee and M. Gerla, "Mobility Prediction and Routing in Ad hoc Wireless Networks," *International Journal of Network Management*, vol. 11, no. 1, pp. 3-30, 2001.
[19] Z. Ye, S. V. Krishnamurthy and S. K. Tripathi, "A Framework for Reliable Routing in Mobile Ad Hoc Networks," *Proceedings of the IEEE International Conference on Computer Communications*, 2003.
[20] X. Zeng, R. Bagrodia and M. Gerla, "GloMoSim: A Library for Parallel Simulation of Large-Scale Wireless Networks," *Proceedings of the 12th Workshop on Parallel and Distributed Simulations*, Banff, Alberta, Canada, May 26-29, 1998.

Optimal Allocation of Rates in Guaranteed Service Networks

Aniruddha Diwan
Motorola India Electronics Ltd
Bagmane Techpark, C. V. Raman Nagar, Bangalore 560093, India
E-mail: aniruddha@motorola.com

Joy Kuri
Department of Electronic Systems Engineering, Indian Institute of Science, Bangalore 560012, India
E-mail: kuri@cedt.iisc.ernet.in, http://www.cedt.iisc.ernet.in/people/kuri/

Sugata Sanyal
School of Technology & Computer Science
Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India
E-mail: sanyals@gmail.com, http://www.tifr.res.in/˜sanyal

Keywords: guaranteed service networks, IntServ, optimal rate allocation, rate reservation

Received: October 6, 2011

We examine the problem of rate allocation in Guaranteed Services networks by assigning a cost corresponding to a rate, and examining least cost allocations. We show that the common algorithm of allocating the same rate to a connection on all links along its path (called the `Identical Rates` *algorithm henceforth) is, indeed, a least cost allocation in many situations of interest. This finding provides theoretical justification for a commonly adopted strategy, and is a contribution of this paper. However, it may happen that the single rate required is not available on some links of the route. The second contribution of this paper is an explicit expression for the optimal rate vector for the case where* `Identical Rates` *is infeasible. This leads to an algorithm called* `General Rates` *that can admit a connection by allocating possibly different rates on the links along a route. Finally, we simulate the* `General Rates` *algorithm in a dynamic scenario, and observe that it can provide, at best, marginally improved blocking probabilities. Our conclusion is that the performance benefits provided by* `General Rates` *are not compelling enough, and the simpler* `Identical Rates` *suffices in practice.*

Povzetek: V članku je analiziran način alociranja v servisnih omrežjih.

1 Introduction

In this paper we consider the problem of rate allocation in the context of the Guaranteed (G) Service framework [1]. The G Service framework enables the receiver of an individual data flow to compute the rate (say R) that it needs to reserve, so that its QoS requirements are satisfied. Thus, a rate vector $(R, R, \ldots, R)$ is reserved, with the vector having as many elements as there are links on the path between the source and the destination. R is a function of the traffic parameters and QoS requirements of the flow, as well as network characteristics such as the number of hops on the path, the scheduling policy employed at each hop and the propagation delay along each hop.

Now suppose that the rate R is not available on some links of the route. When this happens, the Connection Admission Control (CAC) module is usually programmed to block the incoming flow. But, in fact, sufficient bandwidth may not be available only on a *few* links. There might exist other rate vectors that satisfy the delay constraint and the available link bandwidth constraints. Hence, the CAC module may end up blocking a connection request unnecessarily.

This scenario provides the starting point for the work described in this paper. We are motivated to examine the rate allocation problem in the hope that we may avoid unnecessary connection blocking and thereby achieve a bigger admission region than possible when a single rate (*viz.*, R) is reserved on all nodes along the path. To this end, we assign a *cost* to bandwidth allocation at each hop and look for minimum cost allocations.

Rate allocation in Guaranteed Services networks has a long and rich history [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. However, a common theme that runs through the vast literature is that of *identical* rate assignment at each node. Our approach to this problem is novel because we do not start by making this tacit assumption. We pose and answer the questions:

(a) Are there circumstances in which allocating unequal

rates is beneficial?

(b) What criteria can one use to compare alternate rate allocations? What is the *best* rate allocation that can be done according to a chosen criterion?

To the best of our knowledge, the existing literature does not raise these questions.

We begin by listing in Section 3 a few important results in the areas of arrival and service curves, as well as end-to-end delays in Guaranteed Services. These results are used subsequently in our analysis.

2 Related Work

As we have mentioned above, the question that we address does not seem to have received attention in the literature. In this section, we discuss some papers that analyze questions that are related to, but distinct from, the object of study in our paper.

[4] surveys basic mechanisms to support QoS guarantees. Scheduling and buffer management are the basic techniques to achieve QoS. The authors discuss RSVP, the end-to-end resource reservation protocol that was suggested to ensure delay guarantees in the Internet. It is stated explicitly that resource reservation proceeds on the basis of a *single* reservation rate allocated at each node.

The general class of schedulers called Latency Rate (LR) servers is discussed in [6]. Latency and allocated rate are the two parameters that influence the delay experienced by a packet served by an LR server. To analyze the delay experienced by a connection passing through a network of servers, the authors assume that the *same* rate is allocated to a connection at every node on its path.

[10] studies a general network with multiple multihop connections sharing the nodes in the network, and focuses on obtaining end-to-end delay bounds, as well as understanding issues related to stability. Each session is associated with a minimum backlog clearing rate. With this rate given, the authors obtain formulae yielding end-to-end delay bounds. However, the issue of *how to allocate the service rate at each node* was not addressed; it was assumed that the rates are *given.* In this paper, we are concerned precisely with the question of how to allocate rates so that some pre-defined objective is met. Thus, our work addresses a novel aspect of the problem.

In the same way, [7] allows different allocated rates at different nodes, but the interest is in obtaining end-to-end delay bounds for a *given vector* of rates. On the other hand, our concern is with *how to allocate* the vector of rates, so that target delay guarantees can be provided while keeping aggregate resource consumption (*i.e.,* bandwidth consumption) low. Thus, the problem considered in our work is orthogonal to that considered in this paper.

Yet another paper considering related problems is [11]. Each node on the path of a connection uses a "rate-controlled service discipline," but *traffic shapers* are employed at each hop. Thus, the network model considered in this paper is different from the one in our work, because we do not have traffic shapers before each node.

To summarize, we find two threads in the literature. One group of papers assumes that the *same* rate is allocated at each hop on the path. The second group allows different rates to be allocated at each hop, but the focus is on finding formulae for end-to-end delay; the issue of assessing and comparing different possible rate vector allocations has received hardly any attention. In this paper, our concern is with precisely this question.

3 IntServ and Guaranteed Services

To ensure that the QoS requirements are guaranteed, the G Service requires each node on the route of the connection to dedicate an appropriate rate R and a buffer space B to the flow during its holding time. Also, a data flow is required to declare its flow characteristics at the time of connection setup and each node is required to declare its network characteristics. If the flow is admitted, only those packets of the flow that conform to the characteristics are guaranteed the QoS requirements. The rate R and the buffer space B required to guarantee the QoS requirements are a function of the flow characteristics and the network characteristics. In this paper, we consider only R as the resource requirement.

The flow characteristics are specified in terms of the token bucket parameters which are a triplet $(b, \rho, \hat{p})$, where b is the token bucket size, ρ is the token accumulation rate, and $\hat{p}$ is the peak rate of the flow. Node i specifies its network characteristics in terms of the parameters C_i and D_i; typically, C_i and D_i approximate the departure of the service provided by node i from the "fluid" model of service, in which the flow effectively sees a dedicated channel of bandwidth R between source and receiver.

The arrrival process has an envelope $A^{\text{in}}(\tau)$ that is given by

$$A^{\text{in}}(\tau) = \min\{L + \hat{p}\tau, b + \rho\tau\}, \tau \geq 0 \tag{1}$$

where L represents the packet size. When an identical rate R is reserved at each, the service curve for a tandem of nodes $1, 2, \cdots, N$ is given by [16],

$$S_N(\tau) = \left[\left(\tau - \sum_{i=1}^{N} D_i\right) R - \sum_{i=1}^{N} C_i\right]^+ \tag{2}$$

C_i represents the Maximum Transmission Unit (MTU) on the link leaving the i^{th} node on the path, and D_i is the maximum non-preemption delay at the $i\text{th}$ node. The end-to-end delay bound is the maximum horizontal distance between the arrival curve $A^{\text{in}}(\tau)$ and the service curve $S_N(\tau)$. With this, the following delay bound can be easily obtained[1].

$$D^{bound} = \begin{cases} \dfrac{(b-L^{\max})(\hat{p}-R)}{R(\hat{p}-\rho)} + \dfrac{L^{\max}}{R} + \\ \quad \sum_{i=1}^{N}\left[\dfrac{C_i}{R}+D_i\right],\ \hat{p} > R, \\ \dfrac{L^{\max}}{R} + \sum_{i=1}^{N}\left[\dfrac{C_i}{R}+D_i\right],\ \hat{p} \le R. \end{cases} \quad (3)$$

where $L^{\max}$ denotes the maximum packet size on this connection. If D^{reqd} is the worst-case end-to-end delay that is acceptable for the packets of a flow, $D^{bound} \le D^{reqd}$ gives the minimum rate R that has to be allocated at each node on the route of the flow. We denote this minimum rate by R_{equal}.

Now we describe the delay bound when *different* rates are allowed at the nodes. Let the rate allocated at node i be R_i. The arrival process envelope is $A^{in}(\tau)$ as before. Let $R_{\min} = \min_i R_i$ and $R_{\min} \ge \rho$. It was shown in [12, 6] that the service curve for a tandem of nodes $1, 2, \cdots, N$ is given by (example 3.5 of [12])

$$S_N(\tau) = \left[\left(\tau - \sum_{i=1}^{N} D_i - \sum_{i=1}^{N}\frac{C_i}{R_i}\right)R_{\min}\right]^{+} \quad (4)$$

As before, the end-to-end delay bound is the maximum horizontal distance between the arrival process envelope $A^{in}(\tau)$ and the service curve $S_N(\tau)$. Then the following delay bound can be easily obtained.

$$D^{bound} = \begin{cases} \dfrac{(b-L^{\max})(\hat{p}-R_{\min})}{R_{\min}(\hat{p}-\rho)} + \dfrac{L^{\max}}{R_{\min}} + \\ \quad \sum_{i=1}^{N}\left[\dfrac{C_i}{R_i}+D_i\right],\ \hat{p} > R_{\min}, \\ \dfrac{L^{\max}}{R_{\min}} + \sum_{i=1}^{N}\left[\dfrac{C_i}{R_i}+D_i\right],\ \hat{p} \le R_{\min}. \end{cases} \quad (5)$$

It can be seen that when the rate allocated at all the nodes is identical and R, we recover the delay bound of Eqn. (3).

4 Minimum Cost Rate Allocation

Suppose that a rate R_i is allocated to the flow on link $i, 1 \le i \le N$, in such a way that the QoS requirements of the flow are met. We call $\overline{R} = \{R_1, \cdots, R_N\}$ the "rate vector" assigned to the flow. Then the flow can be admitted if there exists a rate vector assignment $\overline{R}$ that satisfies the following constraints:

1. $R_{\min} = \min_i R_i \ge \rho$, i.e., the minimum allocated rate along the route is greater than the average arrival rate of the flow.
2. $D^{bound} \le D^{reqd}$, i.e., the end-to-end delay seen by packets of the flow is less than the end-to-end delay requirement specified by the flow.
3. $R_i \le \gamma_i, 1 \le i \le N$; γ_i is the available link bandwidth on link i (which may be less than the capacity of the link). This constraint simply says that the allocated rate should not exceed the available link bandwidth.

[1]We assume $R \ge \rho$ because when $R < \rho$, the delay is unbounded.

There may be many rate vector assignments that satisfy the above constraints. We associate a cost with allocating a rate on a link. The cost function $f_i(R_i)$ for link i is assumed to be strictly convex and increasing in the allocated rate R_i. This implies that low-cost rate allocations will tend to avoid assigning large rates. It is further assumed that the cost function is the same for each link and is denoted by $f()$. For a rate vector $\overline{R} = \{R_1, \cdots, R_N\}$, the total cost is defined as $t(\overline{R}) = \sum_{i=1}^{N} f(R_i)$. We would like to allocate *that* rate vector which minimizes the total cost and hence our optimization criterion is minimization of the total cost for the flow.

Without loss of generality, we order the N links in the tandem such that the link with the least available capacity is numbered 1 and so on, i.e., $\gamma_1 \le \gamma_2 \le \cdots \le \gamma_N$. The total cost minimization problem, denoted by **MinimumCost**, is as follows:

$$\textbf{(MinimumCost)} \quad \min \sum_{i=1}^{N} f(R_i)$$

subject to:

$$\frac{(b-L^{\max})(\hat{p}-R_{\min})}{R_{\min}(\hat{p}-\rho)} + \frac{L^{\max}}{R_{\min}} + \sum_{i=1}^{N}\left[\frac{C_i}{R_i}+D_i\right] \le D^{reqd} \quad (6)$$

$$R_i \le \gamma_i, 1 \le i \le N \quad (7)$$

$$R_{\min} \ge \rho \quad (8)$$

We note that *if* $R_1 = R_2 = \cdots = R_N = \rho$ is a feasible solution to the **MinimumCost** problem, it is also the optimal solution. That is, the end-to-end delay requirement is met by simply allocating the average rate of arrivals at each link on the path. To avoid this trivial case, we consider only those connections for which allocating the average rate at each link violates the end-to-end delay requirement. Such connections are called "delay-constrained." In other words, we assume that $(\rho, \rho, \cdots, \rho)$ is *not* a feasible solution to the **MinimumCost** problem.

Substituting $\sigma = (b\hat{p} - \rho L^{\max})/(\hat{p}-\rho)$, we can rewrite the constraint in Eqn. (6) as follows.

$$\frac{\sigma}{R_{\min}} + \sum_{i=1}^{N}\frac{C_i}{R_i} \le D^{reqd} - \sum_{i=1}^{N} D_i + \frac{(b-L^{\max})}{(\hat{p}-\rho)} \quad (9)$$

We denote the R.H.S. of Eqn. (9) by D and note that D is a positive quantity. Let $x_i = 1/R_i, 1 \le i \le N$, $\theta_i = 1/\gamma_i, 1 \le i \le N$ and $\delta = 1/\rho$. Let $x_{\max} = \max_i x_i$. Let $\overline{x} = (x_1, \cdots, x_N)$. Let $h(x_i) = f(1/x_i)$. The two results that follow are stated without proof.

Lemma 1. *If $f(x)$ is a convex non-decreasing (concave non-increasing) function over $x \ge 0$, then $h(x) = f(1/x)$ is a convex non-increasing (concave non-decreasing) function over $x \ge 0$.*

We now recast the **MinimumCost** problem in terms of $h(), x_i, \theta_i$ and δ.

$$\textbf{(MinimumCost)} \quad \min \sum_{i=1}^{N} h(x_i)$$

subject to:

$$\sigma x_{\max} + \sum_{j=1}^{N} C_j x_j \leq D \tag{10}$$

$$x_i \geq \theta_i, \quad 1 \leq i \leq N \tag{11}$$

$$x_i \leq \delta, \quad 1 \leq i \leq N \tag{12}$$

Let $H(\overline{x}) = \sum_{i=1}^{N} h(x_i)$ where $\overline{x} = (x_1, \cdots, x_N)$. The following can be easily proved.

Lemma 2. *$H(\overline{x})$ is a convex function in $\overline{x}$.*

It can be seen that the **MinimumCost** problem has a convex cost function with linear constraints. For such problems, there exist simple necessary and sufficient conditions (in terms of Lagrange multipliers) for a feasible solution to be optimal [17].

5 The Unbounded Link Capacities Case

First, we investigate the special case of the problem where we can allocate any amount of bandwidth on the links, i.e., $\gamma_i = \infty$ and $\theta_i = 0, 1 \leq i \leq N$. We note that the constraint Eqn. (10) is actually equivalent to N constraints:

$$\sigma x_i + \sum_{j=1}^{N} C_j x_j \leq D, \ \ 1 \leq i \leq N \tag{13}$$

We denote by **UnbddRates** this special case, i.e., the **MinimumCost** problem without the available link bandwidth constraints, viz.,

$$\textbf{(UnbddRates)} \quad \min \sum_{i=1}^{N} h(x_i)$$

subject to:

$$\sigma x_i + \sum_{j=1}^{N} C_j x_j \leq D, \quad 1 \leq i \leq N \tag{14}$$

$$x_i > 0, \quad 1 \leq i \leq N \tag{15}$$

$$x_i \leq \delta, \quad 1 \leq i \leq N \tag{16}$$

Now suppose that we decide to allocate an identical rate to the flow on each link along its route. It is easy to compute the minimum identical rate R_{equal} from the delay bound constraint of Eqn. (14); we have $x_{equal} = \dfrac{D}{\sigma + \sum_{i=1}^{N} C_i}$ and $R_{\text{equal}} = \dfrac{\sigma + \sum_{i=1}^{N} C_i}{D}$. Thus $\overline{x}_{\text{equal}} = (x_{\text{equal}}, \cdots, x_{\text{equal}})$ is a feasible solution to the **UnbddRates** problem and among all identical rate vectors, $\overline{R}_{\text{equal}} = (R_{\text{equal}}, \cdots, R_{\text{equal}})$ has the least total cost. This follows because reducing x_{equal} further will only push up the cost $\sum_{i=1}^{N} h(x_i)$, as $h(x_i)$ is a non-increasing function.

The approach of allocating of $\overline{R}_{\text{equal}}$ is widely used to provide end-to-end delay guarantees under the Guaranteed Services framework. We refer to this policy as the `Identical Rates` policy. Our next theorem states that $\overline{x}_{equal}$ need not always be the optimal solution to the **UnbddRates** problem and gives an explicit condition for $\overline{x}_{equal}$ to be the optimal solution.

Theorem 1. *$\overline{x}_{equal}$ is the optimal solution to the* **UnbddRates** *problem* iff $\sigma + \sum_{j=1}^{N} C_j > NC_i, 1 \leq i \leq N$.

Proof: We apply Proposition 3.4.2 of [17]. The constraints of the **UnbddRates** problem are ordered as follows.

- $\sigma x_i + \sum_{j=1}^{N} C_j x_j \leq D$ is the ith constraint, $(1 \leq i \leq N)$.
- $x_i > 0$ is the $(N+i)$th constraint, $(1 \leq i \leq N)$.
- $x_i \leq \delta$ is the $(2N+i)$th constraint, $(1 \leq i \leq N)$.

As seen before, $\overline{x}_{equal}$ is a feasible solution to the **UnbddRates** problem. Let μ_i^* denote the multiplier for the ith constraint.

- If $\overline{x}_{\text{equal}}$ is also the optimal solution, we need to show the existence of appropriate scalars $\mu_i^*, 1 \leq i \leq 3N$ that satisfy the conditions of Proposition 3.4.2 of [17].
- If $\overline{x}_{\text{equal}}$ is *not* the optimal solution, we need to show that there exist no scalars $\mu_i^*, 1 \leq i \leq 3N$ that satisfy the conditions of Proposition 3.4.2 of [17].

Let J and $A()$ be as defined in Proposition 3.4.2 of [17]. Therefore, let $J = \{1, \cdots, 3N\}$ and at $\overline{x}_{equal}$, $A(\overline{x}_{equal}) = \{1, \cdots, N\}$ as these are the only active constraints at $\overline{x}_{equal}$. If $\overline{x}_{equal}$ is the optimal solution to the **UnbddRates** problem, then $\mu_i^* = 0, (N+1) \leq i \leq 3N$ and we need to show that there exist unique nonnegative $\mu_1^*, \cdots, \mu_N^*$, not all zero, such that (using the notation of Proposition 3.4.2 of [17]), $g(x) = f(x) + \sum_{j \in J} \mu_j^* (a_j' x - b_j)$ is minimized at $\overline{x}_{equal}$. In our case,

$$g(\overline{x}) = h(x_1) + \cdots + h(x_n) + \sum_{j=1}^{N} \mu_j^* \left[\sigma x_j + \sum_{i=1}^{N} C_i x_i - D \right] \tag{17}$$

Note that $g(\overline{x})$ is a convex function in $\overline{x}$. If $g(\overline{x})$ is minimized at $\overline{x}_{equal}$, the partial derivatives of $g(\overline{x})$ should vanish at $\overline{x}_{equal}$. If we take partial derivatives w.r.t. $x_i, 1 \leq i \leq N$ and set each partial derivative equal to 0 at $\overline{x}_{equal}$, we get N linear equations for $\mu_j^*, 1 \leq j \leq N$ as follows,

$$\sigma \mu_i^* + C_i \sum_{j=1}^{N} \mu_j^* = -h'(x_{\text{equal}}), 1 \leq i \leq N \tag{18}$$

where $h'(x_{\text{equal}})$ is the derivative of $h(x)$ at x_{equal}. Solving these N equations for μ_i^* gives

$$\mu_i^* = -\frac{h'(x_{\text{equal}})}{\sigma} \frac{\sigma + \sum_{j=1}^{N} C_j - NC_i}{\sigma + \sum_{j=1}^{N} C_j}, i = 1, \cdots, N \tag{19}$$

Note that $h'(x_{equal})$ is a negative quantity. This implies that if $\sigma + \sum_{j=1}^{N} C_j > NC_i \ \forall i$, then $\overline{x}_{equal}$ is the optimal solution. Otherwise, there do not exist non-negative $\mu_i^* \ \forall i \in \{1, \cdots, N\}$ and hence $\overline{x}_{equal}$ is not the optimal solution. ■

There are two points to note here. Firstly, Theorem 1 is true for *any* convex non-decreasing cost function $f()$. Secondly, it can be seen from Theorem 1 that the identical rate vector $\overline{x}_{equal}$ need not always be optimal. We give a numerical example to show that a better rate vector indeed exists if $\sigma + \sum_{j=1}^{N} C_j > NC_i$ is not satisfied for every i.

Let us consider the simple cost function $f(R_i) = R_i$ which means $h(x_i) = 1/x_i$. Thus our optimization criterion is the minimization of the total allocated bandwidth. Consider the following choice of parameters.

- $N = 3$ and $\sigma = 30, C_1 = 25, C_2 = 5, C_3 = 5, \rho = 1, D = 6.5$.

Then it can be seen that $\sigma + \sum_{i=1}^{3} C_i > 3C_2$ and $\sigma + \sum_{i=1}^{3} C_i > 3C_3$, but $\sigma + \sum_{i=1}^{3} C_i < 3C_1$.

- The identical rate vector is $\overline{x}_{equal} = (0.1, 0.1, 0.1)$ for which the total cost is $H(\overline{x}_{\text{equal}}) = 30$.
- Consider $\overline{x}' = (x_1', x_2', x_3')$ where $x_1' = 0.095, x_2' = 0.103 + \dfrac{0.005}{35}, x_3' = 0.103$. It can be easily verified that $\overline{x}'$ is a feasible solution and $H(\overline{x}') = 29.93 < H(\overline{x}_{\text{equal}})$.

This shows that $\overline{x}_{\text{equal}}$ is not the optimal solution as $\overline{x}'$ is a better rate vector.

Corollary 1. *For the case $C_1 = C_2 = \cdots = C_N$ and unbounded rates, $\overline{R}_{\text{equal}}$ is the optimal rate allocation vector.*

C_i corresponds to the MTU at the interface i. When the same link technology is employed in a network backbone (a situation that may occur quite often in practice), we have the special case of equal MTU:$C_1 = C_2 = \cdots = C_N (= C_{\text{equal}})$. In the rest of the paper we work with this assumption.

6 The Bounded Link Capacities Case

Now suppose that the optimal rate R_{equal} is not available on some links of the route. When this happens, the Connection Admission Control (CAC) module is usually programmed to block the incoming flow. But, in fact, sufficient bandwidth might not be available only on a *few* links. There might exist another rate vector which satisfies the delay constraint and the available link bandwidth constraints. We note that if such a rate vector exists, the total cost of this rate vector is greater than $H(\overline{x}_{\text{equal}})$. Thus, this is the price to be paid for admitting the flow: the cost is higher.

When $\overline{x}_{equal}$ is infeasible, there is at least one link which the available bandwidth is less than R_{equal}. By our numbering convention, link 1 has insufficient bandwidth; *i.e.,* $\gamma_1 < R_{\text{equal}}$. Let $\overline{x}^* = (x_1^*, x_2^*, \cdots, x_N^*)$ be the optimal solution when $\overline{x}_{equal}$ is infeasible.

Lemma 3. *When $\overline{x}_{equal}$ is infeasible, the optimal vector $\overline{x}^*$ is characterized by $x_1^* = \theta_1$ and $x_i^* \leq \theta_1, 2 \leq i \leq N$.*

The result above says that when $\overline{x}_{\text{equal}}$ is infeasible, it is optimal to allocate the *entire* available bandwidth on link 1 (reciprocal of θ_1) to the incoming connection. *Proof:* The proof relies on establishing the three claims that follow.

Claim 1: There exists $i \in \{2, \cdots, N\}$ such that $x_i^* < \theta_1$. *Proof:* We have

$$\sigma x_{equal} + \sum_{i=1}^{N} C_{\text{equal}} x_{\text{equal}} = D$$

Let $x_{\max}^* = \max_i x_i^*$. As $\overline{x}^*$ is the optimal solution,

$$\sigma x_{\max}^* + C_{\text{equal}} \sum_{i=1}^{N} x_i^* \leq D$$

But $x_1^* \geq \theta_1 > x_{equal}$. This implies

$$\{\sigma + (N-1)C_{\text{equal}}\} x_{\text{equal}} \geq \sigma x_{\max}^* + C_{\text{equal}} \sum_{i=2}^{N} x_i^*$$

But $x_{\max}^* \geq x_1^* \Rightarrow x_{\max}^* > x_{equal}$. This implies

$$\{\sigma + (N-1)C_{\text{equal}}\} x_{\text{equal}} > \sigma x_{\text{equal}} + C_{\text{equal}} \sum_{i=2}^{N} x_i^*$$

which gives

$$\frac{1}{(N-1)} \sum_{i=2}^{N} x_i^* < x_{equal}$$

Hence there exists an $i \in \{2, \cdots, N\}$ such that $x_i^* < x_{equal} < \theta_1$. Denote it by x_k^*.

Claim 2: $x_1^* = \theta_1$. *Proof:* We prove this by contradiction. Assume that $x_1^* > \theta_1$. Recalling that $x_k^* < \theta_1$, consider a new rate vector $\overline{x}'$ such that, $x_1' = x_1^* - \epsilon > \theta_1$, $x_k' = x_k^* + \epsilon < \theta_1$ and $x_i' = x_i^*, i \neq \{1, k\}$. Let $x_{\max}' = \max_i x_i'$. We choose an ϵ that preserves the structure of $\overline{x}^*$, i.e., if $x_{\max}^* = x_j^*$, then $x_{\max}' = x_j'$. It is easy to see that such an ϵ exists. Note that $\overline{x}'$ satisfies the delay constraint. Also note that

$$h(x_1^*) + h(x_k^*) \geq h(x_1') + h(x_k')$$

as $h()$ is a convex function. This implies $\overline{x}'$ is the optimal rate vector and not $\overline{x}^*$. This is a contradiction. Thus $x_1^* = \theta_1$.

Claim 3: $x_i^* \leq \theta_1, \quad 1 \leq i \leq N$. *Proof:* We know that $x_1^* = \theta_1$ and $x_k^* < \theta_1$. We need to prove the claim for the rest. We do this by contradiction. Let there exist a j (among the rest) such that $x_j^* > \theta_1$. Consider $x_1' = x_1^* + \epsilon$ and $x_j' = x_j^* - \epsilon$. With arguments similar to those of claim 2, we get a better cost at $\overline{x}'$ which contradicts the assumption that $\overline{x}^*$ was the optimum. Thus $x_i^* \leq \theta_1$, as required. ■

6.1 The OneLink Problem

Now we consider the case when $\gamma_1 < R_{equal}$, but there are no restrictions on the available link capacities on other links. We refer to this as the **OneLink** problem and want to obtain the corresponding optimal rate vector. Let $\overline{x}^* = \{x_1^*, \cdots, x_N^*\}$ denote the optimal rate vector. >From Lemma 3, we know that $x_i^* \le \theta_1, i = \{2, \cdots, N\}$ and $x_1^* = \theta_1$. Then the **OneLink** problem is as follows:

$$\textbf{(OneLink)} \quad \min \textstyle\sum_{i=2}^{N} h(x_i)$$

subject to:

$$C_{\text{equal}} \textstyle\sum_{j=2}^{N} x_j \le D - (\sigma + C_{\text{equal}})\theta_1 \qquad (20)$$

$$\sigma x_i + C_{\text{equal}} \textstyle\sum_{j=2}^{N} x_j \le D - C_{\text{equal}}\theta_1, 2 \le i \le N \qquad (21)$$

$$x_i > 0, \quad 2 \le i \le N \qquad (22)$$

$$x_i \le \delta, \quad 2 \le i \le N \qquad (23)$$

Lemma 4. *For the* **OneLink** *problem, the optimal rate vector is such that* $x_2^* = \cdots = x_N^*$.

Proof: By contradiction. Assume there exist $x_j^* \ne x_k^*$. Without loss of generality $x_j^* < x_k^*$. Let $x' = \dfrac{x_j^* + x_k^*}{2}$. Then by convexity of $h()$,

$$h\left(\frac{x_j^* + x_k^*}{2}\right) \le \frac{1}{2}\left\{h(x_j^*) + h(x_k^*)\right\}$$

which gives

$$h(x') + h(x') \le h(x_j^*) + h(x_k^*)$$

contradicting the assumption that $\overline{x}^*$ is the optimal rate vector. ■

It is easy to see that the delay constraint inequality is tight at the optimal rate vector. This gives $x_2^* = \cdots = x_N^* = \dfrac{D - (\sigma + C_{equal})\theta_1}{(N-1)C_{equal}}$. Let $x_{equal}^1 = \dfrac{D - (\sigma + C_{equal})\theta_1}{(N-1)C_{equal}}$ and $R_{\text{equal}}^1 = \dfrac{1}{x_{\text{equal}}^1}$. Let $\overline{x}_{OneLink} = (\theta_1, x_{\text{equal}}^1, \cdots, x_{\text{equal}}^1)$ and $\overline{R}_{OneLink} = (\gamma_1, R_{\text{equal}}^1, \cdots, R_{\text{equal}}^1)$. Summarizing, we have the following:

Theorem 2. *$\overline{x}_{OneLink}$ is the optimal solution to the* **OneLink** *problem and therefore $\overline{R}_{OneLink}$ the optimal rate allocation vector for the* **OneLink** *problem.*

6.2 The TwoLinks Problem

It is possible that $\overline{x}_{equal}$ and $\overline{x}_{OneLink}$ are both not feasible. Then we know that $\theta_1 > x_{equal}$ and $\theta_2 > x_{\text{equal}}^1$. Let $\overline{x}^* = \{x_1^*, \cdots, x_N^*\}$ be the optimal solution for this case.

Lemma 5. *When $\overline{x}_{\text{equal}}$ and $\overline{x}_{OneLink}$ are both infeasible, the optimal rate vector $\overline{x}^*$ is characterized by $x_1^* = \theta_1, x_2^* = \theta_2$ and $x_i^* \le \theta_2, 3 \le i \le N$.*

Proof: From Lemma 3, $x_1^* = \theta_1$ and $x_i^* \le \theta_1$.

Claim 1: There exists $i \in \{3, \cdots, N\}$ such that $x_i^* < \theta_2$. *Proof:* We have

$$\sigma\theta_1 + C_{equal}\theta_1 + \sum_{i=2}^{N} C_{equal} x_{\text{equal}}^1 = D$$

Let $x_{\max}^* = \max_i x_i^*$, then $x_{\max}^* = \theta_1$. As $\overline{x}^*$ is the optimal solution,

$$\sigma x_{\max}^* + C_{equal} \sum_{i=1}^{N} x_i^* \le D$$

Therefore,

$$\sigma\theta_1 + C_{\text{equal}}\theta_1 + C_{\text{equal}} \sum_{i=2}^{N} x_i^* \le D$$

But $x_2^* \ge \theta_2 > x_{\text{equal}}^1$. This implies that

$$(N-2)C_{\text{equal}} x_{\text{equal}}^1 \ge C_{\text{equal}} \sum_{i=3}^{N} x_i^*$$

which gives

$$\frac{1}{(N-2)} \sum_{i=3}^{N} x_i^* < x_{\text{equal}}^1$$

Hence there exists a $i \in \{3, \cdots, N\}$ such that $x_i^* < x_{equal} < \theta_2$. Denote it by x_k^*.

Claim 2: $x_2^* = \theta_2$. *Proof:* Similar to that of Claim 2 of Lemma 3.

Claim 3: $x_i^* \le \theta_2, 3 \le i \le N$. *Proof:* Similar to that of Claim 3 of Lemma 3. ■

Next, we consider the problem where $\gamma_1 < R_{equal}$ and $\gamma_2 < R_{\text{equal}}^1$, but there are no restrictions on the available link capacities on other links. This is the **TwoLinks** problem. Let $\overline{x}^* = \{x_1^*, \cdots, x_N^*\}$ denote the optimal rate vector for this problem. >From Lemma 5, we know that $x_1^* = \theta_1$ and $x_2^* = \theta_2$. Then the problem is:

$$\textbf{(TwoLinks)} \quad \min \textstyle\sum_{i=3}^{N} h(x_i)$$

subject to:

$$C_{\text{equal}} \textstyle\sum_{j=3}^{N} x_j \le D - (\sigma + C_{\text{equal}})\theta_1 - C_{\text{equal}}\theta_2,$$

$$C_{\text{equal}} \textstyle\sum_{j=3}^{N} x_j \le D - C_{\text{equal}}\theta_1 - (\sigma + C_{\text{equal}})\theta_2,$$

$$\sigma x_i + C_{equal} \textstyle\sum_{j=3}^{N} x_j \le D - C_{equal}(\theta_1 + \theta_2)$$

$$3 \le i \le N$$

$$x_i > 0, \quad 3 \le i \le N$$

$$x_i \le \delta, \quad 3 \le i \le N$$

Lemma 6. $x_3^* = \cdots = x_N^*$.

Proof: Same as that of Lemma 4. ■

It is easy to see that the delay constraint inequality is tight at the optimal rate vector. This gives $x_3^* = \cdots = x_N^* = \dfrac{D - (\sigma + C_{equal})\theta_1 - C_{equal}\theta_2}{(N-2)C_{equal}}$.

Let $x^2_{equal} = \dfrac{D - (\sigma + C_{equal})\theta_1 - C_{equal}\theta_2}{(N-2)C_{equal}}$ and
$R^2_{equal} = \dfrac{1}{x^2_{equal}}$.
Let $\overline{x}_{TwoLinks} = (\theta_1, \theta_2, x^2_{equal}, \cdots, x^2_{equal})$ and
$\overline{R}_{TwoLinks} = (\gamma_1, \gamma_2, R^2_{equal}, \cdots, R^2_{equal})$. Then we have:

Theorem 3. *$\overline{x}_{TwoLinks}$ is the optimal solution to the* **TwoLinks** *problem and therefore $\overline{R}_{TwoLinks}$ is the optimal rate allocation vector for the* **TwoLinks** *problem.*

6.3 The K-Links Problem

It is clear that in a similar way, one can have problems where $3, 4, 5, \cdots$ links have limited capacities. The pattern of the optimal solution for the **K-Links**, problem, $K \geq 3$, is already clear from the optimal solutions for the **OneLink** and the **TwoLinks** problems. Let
$R^I_{equal} = \dfrac{(N-I)C_{equal}}{D - \dfrac{\sigma + C_{equal}}{\gamma_1} - \dfrac{C_{equal}}{\gamma_2} - \cdots - \dfrac{C_{equal}}{\gamma_I}}, I \geq 1$ and $x^I_{equal} = 1/R^I_{equal}$.
Let $\overline{R}_{ILinks} = (\gamma_1, \cdots, \gamma_I, R^I_{equal}, \cdots, R^I_{equal})$ and
$\overline{x}_{ILinks} = (\theta_1, \cdots, \theta_I, x^I_{equal}, \cdots, x^I_{equal}), I \geq 1$.
The **K-Links** problem is as follows.

- The available link capacity on link $I, 1 \leq I \leq K$, is γ_I, where $\gamma_I < R^{(I-1)}_{equal}$ and unlimited capacity is available on links $(K+1), \cdots, N$. What is the optimal rate allocation vector that minimizes the total cost and obeys the delay constraint and the available link capacity constraints?

As the technique of obtaining the optimal solution is essentially the same as that in the **TwoLinks** case, we state the results directly, without proof. Let $\overline{x}^* = \{x^*_1, \cdots, x^*_N\}$ be the optimal solution to the **K-Links** problem.

Lemma 7. *$x^*_i = \theta_i, 1 \leq i \leq K$ and $x^*_i \leq \theta_K, (K+1) \leq i \leq N$.*

Lemma 8. *$x^*_{(K+1)} = \cdots = x^*_N$.*

Theorem 4. *$\overline{x}_{KLinks}$ is the optimal solution to the* **K-Links** *problem and therefore $\overline{R}_{KLinks}$ is the optimal rate allocation vector for the* **K-Links** *problem.*

In practice, the available capacities of the links are always bounded. The previous results lead directly to a `General Rates` allocation algorithm that can be used to decide whether a flow can be admitted and, if so, the optimal rate allocation vector.

7 Simulation Results

`General Rates` allows an incoming connection to be possibly accepted even when `Identical Rates` blocks it. But, as we have remarked earlier, the total bandwidth required to be allocated is more for `General Rates`. This means that less bandwidth may be available for calls in the future. So even though `General Rates` looks appealing in the short term, the long-term consequences of following it need to be examined. In this section, we simulate the `General Rates` algorithm to study this aspect.

We provide some details of our simulation scenario. At each link, the packets are scheduled according to the Weighted Fair Queueing (WFQ) policy [18, 9, 10]. WFQ falls under the Guaranteed Services framework with the following parameters: For a flow j at link i, $C_i = L^{\max}_j$ and $D_i = \dfrac{L^{\max}}{\gamma_i} + d^{\text{prop}}_i$, where $L^{\max}_j$ is the maximum packet size of flow j, $L^{\max}$ is the maximum size of the packets at link i, γ_i is the capacity of link i and d^{prop}_i is the propagation delay of link i.

We assume that all the connections to be routed are full-duplex, that all links are bidirectional and the two halves of a full-duplex connection are to be routed on the same path. Connection requests are assumed to arrive according to a Poisson process and last for a duration that is exponentially distributed. We further assume that the traffic specifications and delay requirements of the connection requests are as specified in Table 1. Our performance criterion is the blocking probability of connections.

Class	b(kB)	ρ(Mbps)	$\hat{p}$(Mbps)	D^{reqd}(ms)	$L^{\max}$(kB)
Voice	0.1	0.064	0.064	50	0.1
Vid conf	10	0.5	10	75	1.5
Stored vid	100	3	10	100	1.5

Table 1: Traffic Parameters

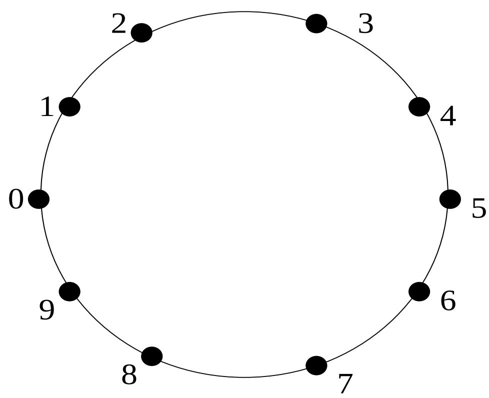

Figure 1: The homogeneous ring topology, with 155 Mbps links and propagation delay of 4 ms on each link.

To study the impact of the `General Rates` policy, we consider source-destination (SD) pairs with at least 2 hops on the path. We consider the homogeneous ring topology of Fig 1 and the NSFNET topology of Fig. 2. Both 2 hops and 3 hops away SD pairs are considered.

To study the effect of non-identical link characteristics, we also consider the ring and NSFNET topologies with link

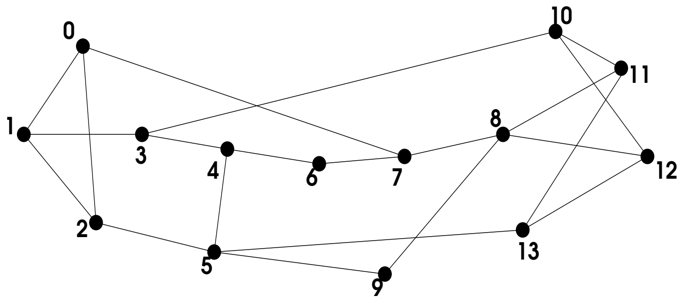

Figure 2: The NSFNET topology, with 155 Mbps links and propagation delay of 4 ms on each link.

link	rate (Mbps)	delay (ms)
0 1	155	4
0 9	310	4
1 2	155	4
2 3	310	4
3 4	155	4
4 5	310	4
5 6	155	4
6 7	310	4
7 8	155	4
8 9	310	4

Table 2: Link parameters for the nonhomogeneous ring topology.

link	rate (Mbps)	delay (ms)
0 1	155	4
0 2	310	4
0 7	620	4
1 2	155	4
1 3	310	4
2 5	620	4
3 4	155	4
3 10	310	4
4 5	620	4
4 6	310	4
5 9	620	4
5 13	155	4
6 7	310	4
7 8	620	4
8 9	155	4
8 11	310	4
8 12	620	4
10 11	155	4
10 12	310	4
11 13	620	4
12 13	155	4

Table 3: Link parameters for the nonhomogeneous NSFNET topology.

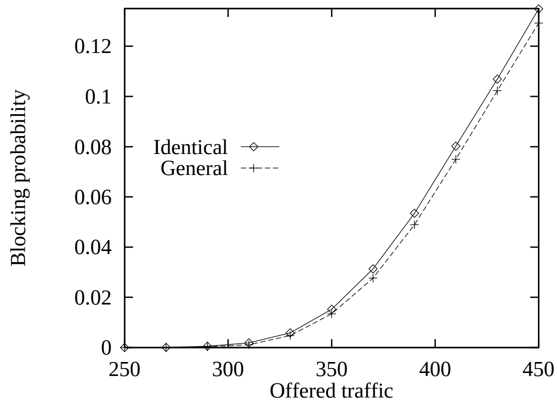

Figure 3: Blocking probabilities for the ring topology of Fig. 1; uniform load, 2 hops away SD pairs.

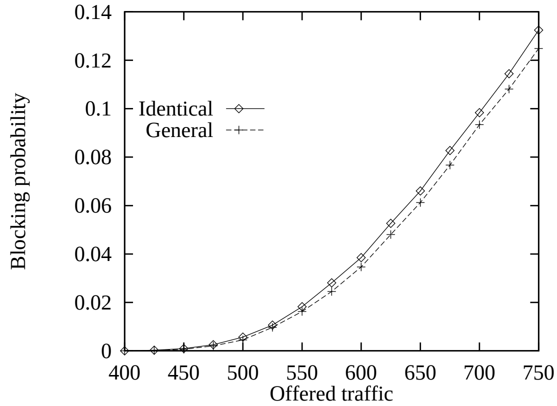

Figure 4: Blocking probabilities for the NSFNET topology of Fig. 2; uniform traffic distribution, 2 hops away SD pairs.

bandwidths taking values 155 and 310 Mbps, and 155, 310 and 620 Mbps, respectively, as shown in Tables 2 and 3.

7.1 Uniform Load

First, we consider the situation when traffic load is *uniform* across all SD pairs. In the next subsection, we present results when traffic load is distributed among SD pairs in an uneven manner.

In Figs 3 to 6, we consider topologies that are "homogeneous" in the sense that link capacities are equal. In Figs 7 to 9, link capacities are unequal. Each of the figures shows a plot of connection blocking probability as traffic load increases.

Fig 3 shows that for the homogeneous ring topology and 2-hop away SD pairs, the `General Rates` policy achieves slightly better (lower) connection blocking probability than the `Identical Rates` policy. The same is true for the homogeneous NSFNET topology (Fig 4). However, when we consider 3-hop away SD pairs, the General Rate policy is unable to show any appreciable improvement (Figs 5 and 6).

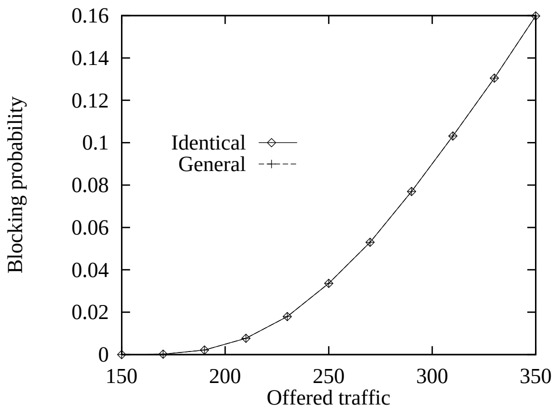

Figure 5: Blocking probabilities for the NSFNET topology of Fig. 2; uniform traffic distribution, 3 hops away SD pairs.

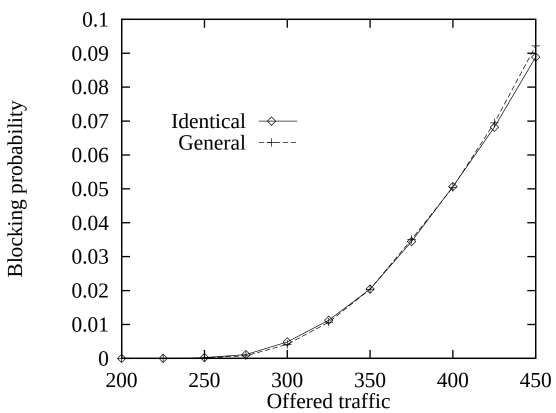

Figure 6: Blocking probabilities for the NSFNET topology of Fig. 2; uniform traffic distribution, 2 & 3 hops away SD pairs.

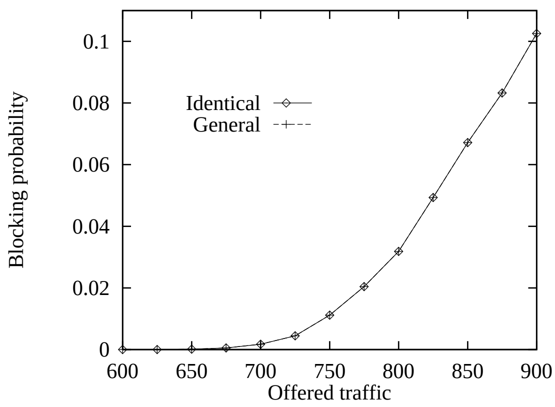

Figure 7: Blocking probabilities for the ring topology of Fig. 1 with link parameters of Table 2; uniform load, 2 hops away SD pairs.

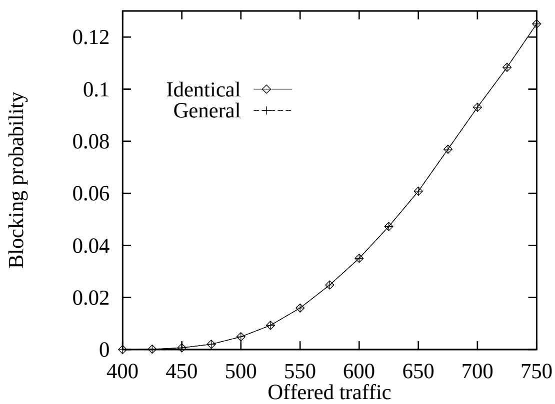

Figure 8: Blocking probabilities for the NSFNET topology of Fig. 2 with non-identical link parameters of Table 3; uniform traffic distribution, 2 hops away SD pairs.

Next, we consider nonhomogeneous topologies. Figs 7, 8 and 9 show that once again, the `General Rates` and `Identical Rates` policies perform similarly as far as connection blocking probability is concerned.

7.2 Nonuniform Load

With nonuniform loads, the general trends are similar. Figs 10 and 11 show that for the homogeneous case (link capacities equal), as long as the SD pairs are 2 hops away, the `General Rates` policy shows a slight improvement in connection blocking. However, as we include SD pairs that are 3 hops away, the advantage disappears (Fig 12 and Fig 13).

Similarly, for nonhomogeneous topologies, there is little to distinguish between the performances of the `General Rates` and the `Identical Rates` policies. We show the 2-hop away SD pair case for the Ring topology in Fig 14, the 2-hop away SD pair case for the NSFNET topology in Fig 15. The case with a mixture of 2-hop away and 3-hop away SD pairs is shown in Fig 16.

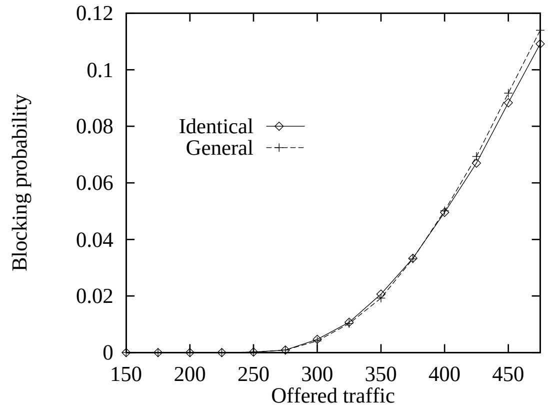

Figure 9: Blocking probabilities for the NSFNET topology of Fig. 2 with non-identical link parameters Table 3; uniform traffic distribution, 2 & 3 hops away SD pairs.

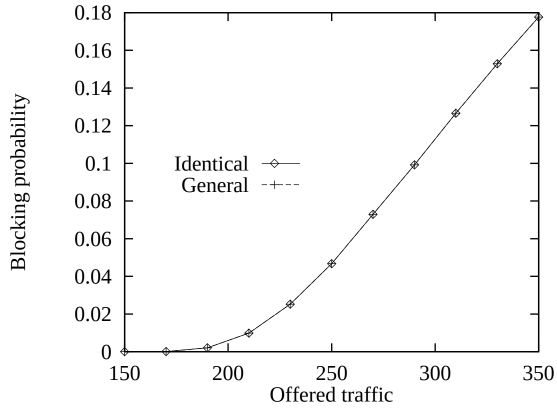

Figure 12: Blocking probabilities for the NSFNET topology of Fig. 2; nonuniform load, 3 hops away SD pairs.

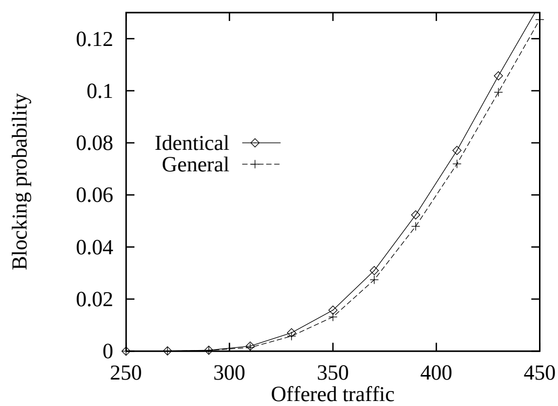

Figure 10: Blocking probabilities for the ring topology of Fig. 1; nonuniform load, 2 hops away SD pairs.

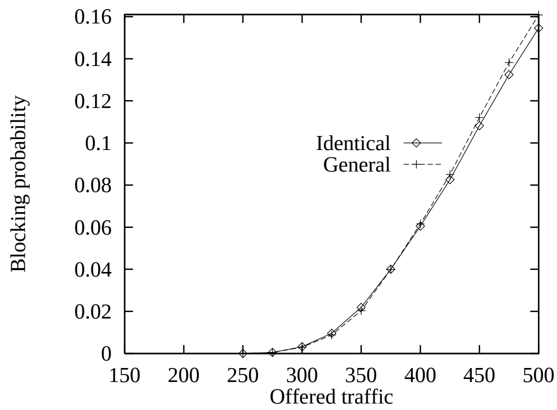

Figure 13: Blocking probabilities for the NSFNET topology of Fig. 2; nonuniform load, 2 & 3 hops away SD pairs.

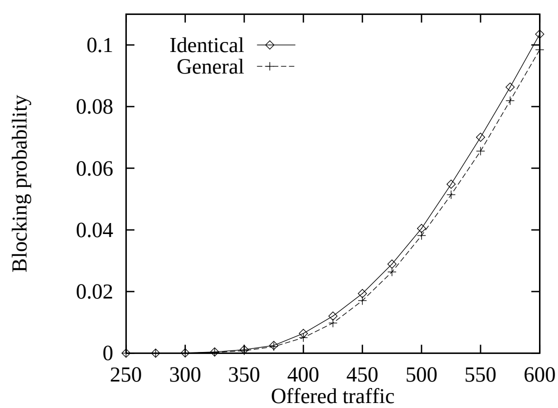

Figure 11: Blocking probabilities for the NSFNET topology of Fig. 2; nonuniform load, 2 hops away SD pairs.

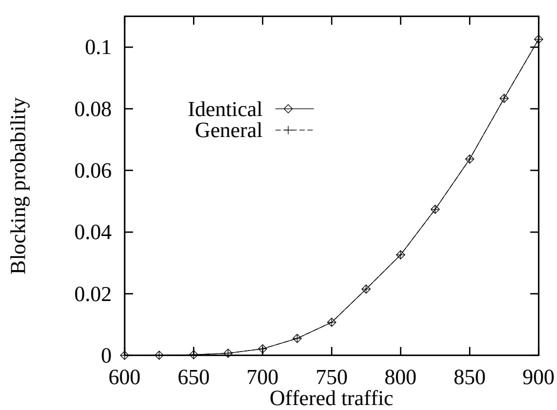

Figure 14: Blocking probabilities for the ring topology of Fig. 1 with link parameters of Table 2; nonuniform load, 2 hops away SD pairs.

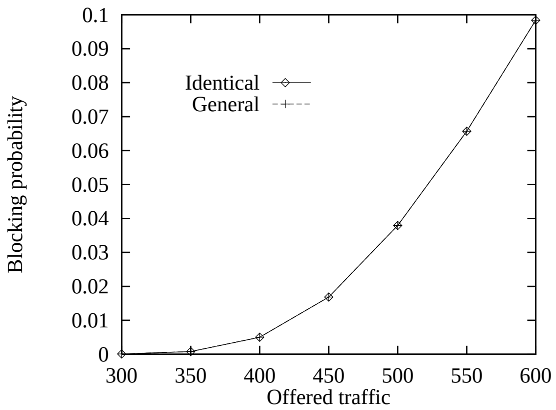

Figure 15: Blocking probabilities for the NSFNET topology of Fig. 2 with link parameters of Table 3; nonuniform load, 2 hops away SD pairs.

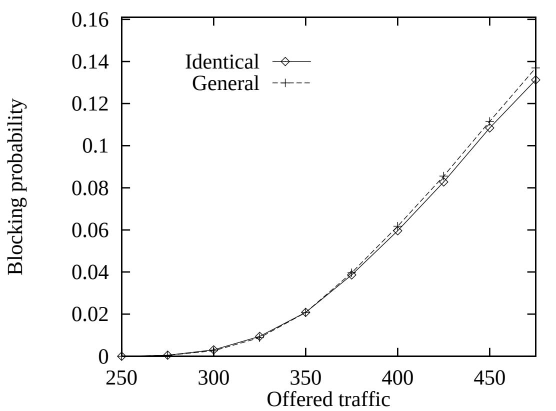

Figure 16: Blocking probabilities for the NSFNET topology of Fig. 2 with link parameters of Table 3; nonuniform load, 2 & 3 hops away SD pairs.

7.3 Summary

The series of plots indicates that the blocking performances achieved by the `Identical` and `General` policies remain close, with slight improvement in one or the other in some cases. The results indicate that there is not much to choose between the two policies as far as blocking performance is concerned.

8 Connection Blocking

We saw in the previous section that accepting a connection whenever possible (what the `General Rates` policy does) may lead to significant blocking of future connections. In this section, we attempt to obtain some insight into this aspect by using very simple arguments.

Suppose that we have a very simple "network" with two links in tandem. Let the capacities of the links be 2.5 and 10 units respectively. Type A connections traverse both the links while type B connections pass over the second link only. The average rate of traffic from a connection of either type is 0.45. The delay requirement of a type A connection can be met by allocating 2 units of bandwidth on the two links. Similarly, the delay requirement of the Type B connection can be met by allocating 2 units on the second link.

For the network and traffic types above, the `Identical Rates` policy would admit only one Type A connection. Now let us consider the `General Rates` policy. Let us assume that it is possible to accommodate a *second* Type A connection by allocating 0.5 and 6 units on the first and second links, respectively. Let us now suppose that the network is empty to begin with, and two Type A connections arrive in quick succession. Also, we assume that these connections have infinite lifetimes.

The `Identical Rates` policy will admit the first Type A connection and block the second, as well as all future Type A connections. The fraction of Type B connections that will be blocked can now be obtained using the Erlang-B formula. The bandwidth available on the second link is 8 units and, therefore, 4 Type B connections can be accommodated; the $M/M/4$ model applies.

The `General Rates` policy will admit both the Type A connections. As a result, the bandwidth available on the second link is only 2 units. For this case, the $M/M/1$ model applies. Clearly, the overall blocking probability in this case will be much higher: the fraction of Type A connections blocked is the same as for `Identical Rates`, while the fraction of Type B connections blocked is much more.

Admittedly, our example network and traffic scenario are very simple and contrived. Nevertheless, they do highlight the problem that can result when the `General Rates` policy is followed. We are currently working on a complete analysis of this problem to extend the intuitive understanding that can be obtained from this simple model.

9 Conclusions

In this work, we study the problem of rate allocation from the perspective of total cost minimization. It is shown that the `Identical Rates` policy need not always minimize the total cost, and a specific condition is derived which, if satisfied, makes the `Identical Rates` policy optimal. However, if the computed identical rate is not available on every link along the path of a connection, `Identical Rates` blocks a request without searching for other rate allocations. To admit a connection whenever it is possible at all, a `General Rates` algorithm is developed; it computes an optimal rate vector with possibly different rates allocated on different links. However, `Genenral Rates` is forced to allocate more bandwidth on the end-to-end path, and this leaves less resources for future connections. Simulations show that the increased bandwidth consumption of `General Rates` can become significant in the long run, as a result of which the aloorithm is unable to show appreciably improved blocking probability. Hence, the simpler `Identical Rates` algorithm is sufficient for all practical purposes.

References

[1] S. Shenker, C. Partridge, and R. Guérin. Specification of guaranteed quality of service. RFC 2212, September 1997.

[2] R. Nagarajan, J. Kurose, and D. Towsley. Allocation of local Quality of Service constraints to meet end-to-end requirements. In *IFIP Workshop on the Performance Analysis of ATM Systems*, Martinique, January 1993.

[3] D. Raz and Y. Shavitt. Optimal partition of QoS requirements with discrete cost functions. In *IEEE INFOCOM*, 2000.

[4] R. Guérin and V. Peris. Quality-of-Service in packet networks: basic mechanisms and directions. *Computer Networks*, 31(3):169–179, February 1999.

[5] P. White. RSVP and integrated services in the Internet. *IEEE Communications Magazine*, May 1997.

[6] D. Stiliadis and A. Varma. Latency-rate servers: a general model for analysis of traffic scheduling algorithms. *IEEE/ACM Transactions on Networking*, 6(5):611–624, October 1998.

[7] P. Goyal and H. Vin. Generalized guaranteed rate scheduling algorithms: a framework. *IEEE/ACM Transactions on Networking*, 5(4):561–571, August 1997.

[8] H. Zhang. Service disciplines for guaranteed performance service in packet-switching networks. *Proceedings of the IEEE*, pages 1374–1396, October 1995.

[9] A. Parekh and R. Gallagher. A generalized processor sharing approach to flow control in integrated services networks: the single node case. *IEEE/ACM Transactions on Networking*, 1(3):137–150, 1993.

[10] A. Parekh and R. Gallagher. A generalized processor sharing approach to flow control in integrated services networks: the multiple node case. *IEEE/ACM Transactions on Networking*, 2(2):347–357, 1993.

[11] V. Peris L. Georgiadis, R. Guérin and K. Sivarajan. Efficient network QoS provisioning based on per node traffic shaping. *IEEE/ACM Transactions on Networking*, 4(4):482–501, August 1996.

[12] R. Agrawal and R. Rajan. Performance bound for guaranteed and adaptive services. Technical Report RC20649, IBM Research Report, December 1996.

[13] L. Georgiadis, R. Guérin, and A. Parekh. Optimal multiplexing on a single link: delay and buffer requirements. In *IEEE INFOCOM*, 1994.

[14] S. Shenker and L. Breslau. Two issues in reservation establishment. In *SIGCOMM Symposium on Communications Architectures and Protocols*, Cambridge, Massachusetts, September 1995.

[15] H. Ahmadi, J. Chen, and R. Guérin. Dynamic routing and call control in high-speed integrated networks. In 13*th International Teletraffic Congress*, pages 19–26, Copenhagen, Denmark, 1991.

[16] L. Georgiadis, R. Guérin, V. Peris, and R. Rajan. Efficient support of delay and rate guarantees in an Internet. In *ACM SIGCOMM*, pages 106–116, August 1996.

[17] D. Bertsekas. *Nonlinear Programming*. Athena Scietific, Belmound, Massachusetts, 1995.

[18] A. Demers, S. Keshav, and S. Shenker. Analysis and simulation of a fair queueing algorithm. In *ACM SIGCOMM*, pages 3–12, 1989.

Issues in Energy Optimization of Reinforcement Learning Based Routing Algorithm Applied to Ad-hoc Networks

Shrirang Ambaji Kulkarni
Department of Computer Science and Engineering
Gogte Institute of Technology, Belgaum – 590006, Karnataka, India
E-mail:sakulkarni@git.edu

G. Raghavendra Rao
Department of Computer Science and Engineering
National Institute of Engineering, Mysore - 590008, Karnataka, India
E-mail: grrao56@gmail.com

Keywords: ad-hoc networks, routing protocols, reinforcement learning, energy optimization and scalability

Received: January 5, 2012

Ad-hoc networks represent a class of networks which are highly unpredictable. The critical work of such networks is performed by the underlying routing protocols. Decision in such an unpredictable environment and with a greater degree of successes can be best modelled by a reinforcement learning algorithm. In this paper we consider SAMPLE, a collaborative reinforcement learning based routing algorithm, which performs competitively with other routing protocols of similar category. A major concern of SAMPLE is its energy consumption, as most of the wireless nodes are driven by finite battery power. Energy conservation has a direct bearing on the network survivability and also affects the underlying quality of services. Energy conservation is not just the problem of the network layer, but it must be considered at the data link layer. Thus we consider a cross-layer energy conservation algorithm SPAN which models a solution by aiding the routing protocol at the network layer with a backbone of stable energy nodes and conserves the energy of remaining nodes by extracting the best features of IEEE 802.11 power saving mode. Most of the network survivability issues should consider scalable scenarios and our work extends our applied energy optimization framework to scalable scenarios. A scalable scenario can be best visualized if we can gain insight into the performances of underlying mobility models. Thus we extend our work to the analysis of the underlying mobility model and its impact on energy conservation in the traffic-mobility dimensions. We also verify the simulation results statistically using hypothesis testing to prove the superiority of our energy conservation attempts for SAMPLE.

Povzetek: Opisana je optimizacija potrošnje s spodbujevalnim učenjem v omrežjih.

1 Introduction

Ad-hoc networks represent a special class of dynamic networks. An ad-hoc network is characterized by dynamic topology, congestion, energy and security constraints [1]. A basic requirement of an ad-hoc network is that nodes act as both hosts and routers. A major work of such networks which represent a complex optimization problem is performed by the underlying routing protocols. Reinforcement learning algorithms hold a lot of promise where solutions to problems in unpredictable scenarios need to be considered [2], [3]. Thus we consider SAMPLE [4] a collaborative reinforcement learning (CRL) routing algorithm which models the dynamics of ad-hoc networks as a distributed optimization problem and strives to solve it, in an optimal manner. Energy consumption modelling was not measured in SAMPLE and in [5] AODV performed better than SAMPLE by 34%. Thus we propose to modify SAMPLE and integrate it with a cross-layer energy extension SPAN [6]. SPAN is engaged in helping a routing algorithm by providing a backbone of energy stable nodes called as coordinators and at the same time enables the remaining nodes to exploit IEEE power saving mode (PSM) [7] of 802.11 at data link layer. Scalability of nodes has a profound effect on the performance of routing algorithms and many hidden issues may come to light. As such the present study extends the energy consumption analysis of SAMPLE under moderately scalable scenarios. To gain some insight into the mobility nodes and their performances the underlying Steady State Random Waypoint [8] for performance analysis is analysed. We also conduct t-Test to vary our assumption and prove the statistical significance of our results. Thus the goal of this paper is to optimize the energy conservation of SAMPLE by integrating it with SPAN and comparing its performance with AODV, a benchmark routing protocol, in the traffic-mobility dimensions and under moderately scalable conditions.

In Section 2 we discuss some of the related work. Section 3 proposes our optimization framework for energy conservation of SAMPLE routing algorithm. In Section 4 we discuss about the Steady State Random Waypoint Mobility Model with its performances under varying density of nodes. Section 5 brings out the simulation results of energy conservation of SPAN with SAMPLE in comparison with a benchmark ad-hoc routing protocol AODV under the traffic-mobility dimensions and scalable conditions along with statistical analysis of results. Section 6 concludes the paper.

2 Related Work

Dowling et.al [4] introduced and evaluated collaborative reinforcement learning (CRL) as a self-organizing technique for mobile ad-hoc network routing protocol called SAMPLE. They considered the Random Waypoint Mobility Model and did not model nor optimize energy conservation as a part of their work. P Nurmi [14] applied reinforcement learning algorithms for ad-hoc network problem where routing decisions of nodes where modelled on selfishness and on the energy level of nodes. Chen and Chang [15] in their work on impact of mobility models on energy conservation showed significant energy conservation differences on various mobility models. They concluded that reactive protocols perform well when nodes move in groups, in terms of energy conservation. They did not evaluate the performance of mobility models in terms of their mobility metrics and did not correlate this to energy conservation. S.A. Kulkarni and G R Rao [25] in their work on energy efficiency of routing protocols observed that DSR was energy efficient as compared to AODV.

They enhanced the energy efficiency of AODV by deploying SPAN, which acted as a middleware between the data link and the network layer. V Naumov and T Gross [16] analysed the scalability of routing methods for ad-hoc networks and investigated the performance of two routing protocols namely AODV and DSR. They considered only Random Waypoint Mobility Model. They did not take energy conservation of routing protocols in their study. Xu et.al. [17] proposed GAF an energy conservation scheme similar to SPAN. GAF had differences as compared to SPAN; firstly it required nodes to know their geographic positions and second was SPAN's superiority in terms of its integration with IEEE 802.11 PSM. In PAMAS the power-saving medium access protocol [18], [19], a node turns off its radio when it is overhearing a packet not addressed to it. This approach justified the idea when it is costly to process a received packet as compared to mere listening.

3 Energy Optimization Framework

In order to optimize energy conservation in the complex dynamics of ad-hoc networks, and support the functioning of the collaborative reinforcement learning algorithm SAMPLE [4], the following architecture illustrated in Figure 1 is proposed.

SAMPLE routing protocol works in an on-demand fashion based on collaborative reinforcement learning (CRL). Routing information is represented as a V-value and is floated in the network by attaching it to the data packets. Each agent maintains routing tables that stores the cost to the neighbours and their last advertised V-value. The path selection is based on the Boltzmann-action selection and optimal paths are selected. Now the

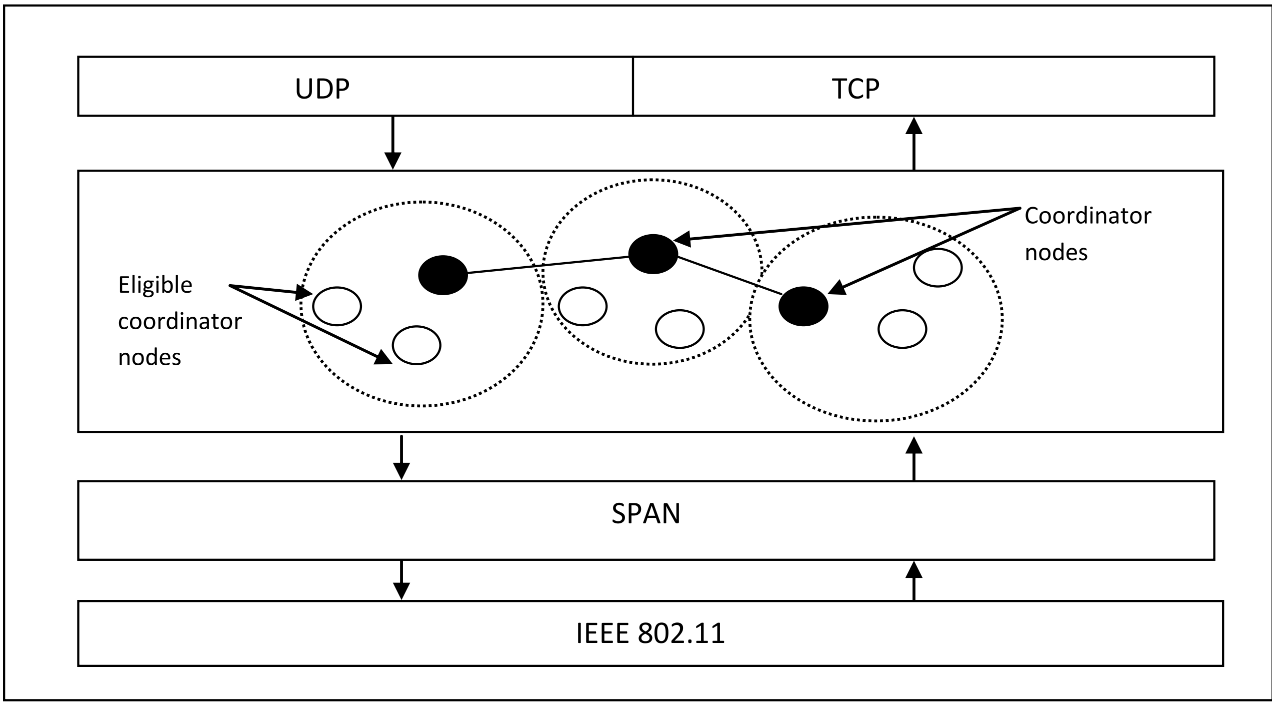

Figure 1: Energy Optimization Frame work for SAMPLE.

active nodes performing critical routing actions are eligible to act as coordinators to form energy-stable backbone nodes as chosen by the SPAN algorithm. SPAN adaptively elects "coordinators" from the given active nodes in the ad-hoc network. The chief role of the coordinators is to stay awake and perform multi-hop routing functions, while other nodes are engaged in power saving mode. The task of the coordinator is rotated among all the nodes and SPAN strives to minimize the number of coordinators at any given point of time, one per group of nodes within a radio range. In order to prevent a situation where several nodes contest simultaneously to be elected as a coordinator, a node delays its candidature as specified in Equation 1.

$$d = \left(\left(1 - {}^{Er}/_{Em}\right) + \left(1 - {}^{Ci}/_{\binom{Ni}{2}} + R\right) \right) * Ni * T \qquad (1)$$

where Ni is the number of neighbours of node i; Ci is the number of additional pairs of nodes that would be connected if "i" becomes coordinator. Er and Em are the amounts of remaining and maximum energies of the node. T is the round-trip delay of a packet and R is a uniform value from the interval (0, 1).

Only if a node satisfies Equation1, does it announce itself as a coordinator by broadcasting a "Hello" packet. SPAN in order to save energy consumption, relies heavily on IEEE 802.11 power saving mode. The basic idea behind PSM is to power down or make the radio device sleep, when it has no data to transmit or receive. Energy conservation study [6] did not consider scalability issues. In our work we evaluate our optimization framework under scalable routing scenarios. One of the important scenarios to gain information about energy conservation under scalable operations is to gain insights in the underlying mobility model and explore its optimization features via its performance analysis metrics. This study substantiates the fact that network lifetime is defined by the time to a partition in the network [9].

4 Steady State Random Waypoint Mobility Model

The Random Waypoint Mobility despite being simple and easy to simulate is not very realistic. The model also suffers from non steady-state distribution at the start of a simulation and then ultimately converges to a steady-state distribution known in probability terms as "stationary distribution". The network performance may be affected between start-up time and after steady state has been reached as described in [26]. Thus we consider a Steady State Random Waypoint Mobility Model.

Consider a mobile node [27] that lives in a connected set A. The trip end times are $T_0 = 0 < T_1 < T_3 < \ldots$ Let $S_n = T_{n+1} - T_n$ be the duration of the n^{th} trip. At trip transition time T_n, the node selects the path $P_n : [0,1] \rightarrow A$ of the n^{th} trip. The mobile position at time t is given by Equation 2.

$$X(t) = Pn\left(\frac{t - T_n}{S_n}\right), \qquad T_n \leq t\, T_{n+1} \qquad (2)$$

The node position at the trip begin-point *is* $X(T_n) = P_n(0)$ and the trip end-point $X(T_{n+1}) = P_n\,(1)$. Denote $U(t) = {}^{t - T_n}/_{S_n}$ the fraction of the elapsed time on the nth trip. Thus, we can write $X(t) = P(t)U(t)$ where $P(t) = P_n,\ T_n \leq t \leq T_{n+1}$. Also assume that the trip selection rule is Markov modulated. The trip selection rule depends on all past only through the current mobile location M_n and the state of a Markov chain I_n. Further, I_n depends on all past only through the last state I_{n-1}. More precisely, I_n (the phase) is defined on some enumerable set I; the sequence of phases (I_n) take values on state space I. Trip selection rule then states that at transition instant Tn, the path Pn and the trip duration S_n, given the phase I_n and mobile position $M_n = X\,(T_n)$ are drawn independently of $"n"$ and any other past. In simple cases, the phase corresponds to being in a pause or move state. Further, the perfect sampling algorithm implemented [10] does not require knowledge of generic constants such as average distance between two random points on a graph which may result in difficulty in computation.

Thus, in the case of Steady State Random Waypoint Model at a trip transition instant, a node picks a trip destination randomly and uniformly on a rectangular area and samples the numeric speed from a uniform distribution. After reaching the trip destination, the node pauses for some duration drawn from a uniform distribution. This process repeats for every trip selection rule.

4.1 Performance Evaluation of Steady State Random Waypoint Mobility Model

The present study evaluated the performance of Steady State Random Waypoint model as specified [11],[12] to gain an insight into scalability issues and consider the performance of Steady State Random Waypoint Model. The study proves that the performance of mobility models has a direct bearing on the performance of routing protocols and this in turn affects the energy conservation capabilities. Around 20 mobile nodes for a low density mode (LD) and around 100 nodes for a medium density mode (MD) are taken into account. This is moderately scalable in a region of 1200 x 1200 m and with speed varying from 5,10,15,20 and 25 m/sec.

4.2 Mobility Performance Analysis Metric

The mobility analysis metrics from [11] were considered for the analysis of mobility models.

- **Average Link Duration**: - This metric considers the interval of longest time [t1, t2] for nodes i and j for the link (i, j). This is then averaged for all node pairs for all existing links specifying the Equation 3.

$$LD' = \frac{\sum_{t1=0}^{T}\sum_{i=1}^{N}\sum_{j=i+1}^{N} LD(i,j,t1)}{P} \quad (3)$$

where P is no of tuples (i, j, t1) and LD (i, j, t1) $\neq 0$

- **Average Relative Speed**:- Relative speed is given by Equation 4.

RST (i, j, t) = |Vi(t) – Vj(t) | (4)

where Vi(t) and Vj(t) is the velocity vector of node i and j at time t. The average value of RST(i, j, t) is given by Equation 5.

$$\text{RST}' = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} RST(i,j,t)}{P} \quad (5)$$

where P is no of tuples (i, j, t1) and RST′(i, j, t1) $\neq 0$

- **Average degree of spatial dependence**: - It specifies the amount of similarity of velocities of given two nodes, given by Ds(i, j, t) and averaged over pair of nodes and time instants and specified by the Equation 6.

$$Ds' = \frac{\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=i+1}^{N} Ds(i,j,t1)}{P} \quad (6)$$

where P is no of tuples (i, j, t1) and Ds′(i, j, t1) $\neq 0$

4.3 Mobility Performance Analysis Results

The Steady State Random Waypoint Mobility Model is analysed for low density (RW-LD) of mobile node and for medium density of mobile nodes (RW-MD) and the results is illustrated in Figure 2, Figure 3 and Figure 4.

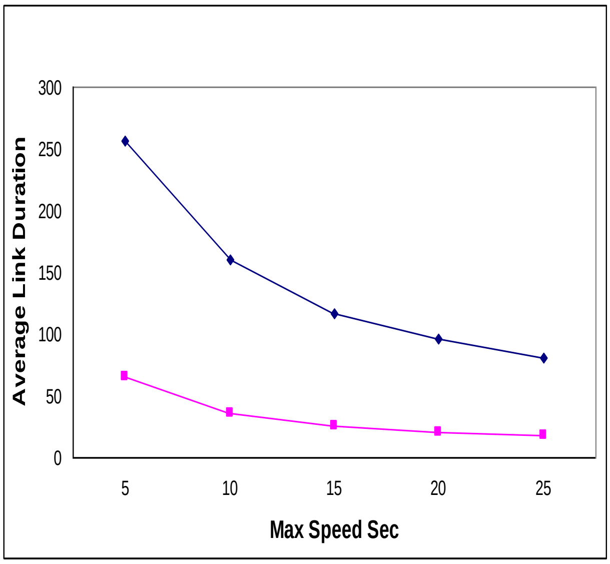

Figure 2: Average link duration of Steady State Random Waypoint Model.

In Figure 2 it is seen that the average link duration, which indicates the stability of the links, is higher for Steady State Random Waypoint Model with low density of nodes (RW-LD) and performs poorly for Steady State Random Waypoint Mobility Model with medium density of nodes (RW-MD). RW-LD in terms of link stability outperforms RW-MD on an average by 76.95 %.

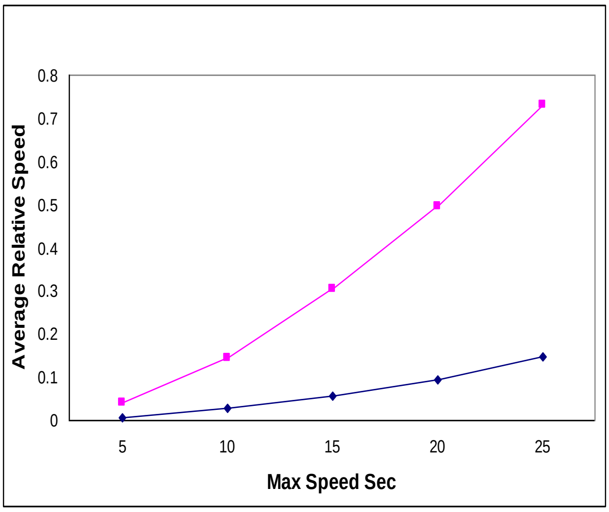

Figure 3: Average Relative Speed of Steady State Random Waypoint Mobility Model.

In Figure 3 it can be observed that the average relative speed is high for RW-MD as compare to RW-LD and this in turn affects the link durations. Thus the high speed dynamics of RW-MD has an effect on the stability of links.

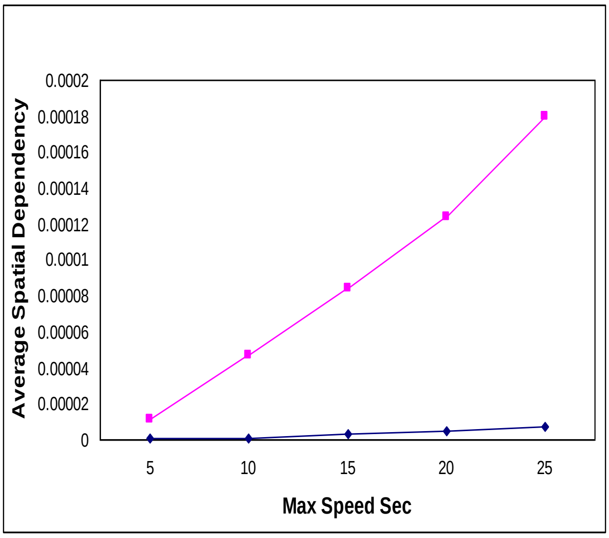

Figure 4: Average Spatial Dependency of Steady State Random Waypoint Mobility Model.

Figure 4 shows that the average spatial dependency of RW-MD is high as compared to RW-LD and RW-MD outscore RW-LD by 96%. As mobile ad-hoc networks exhibit cohesive properties, this is directly supported by high spatial dependency. Also a high spatial dependency minimizes network partitions and routing overhead sometimes as nodes are nearby. This in turns has a positive effect on energy conservation, provided the link duration is also correspondingly high.

5 Simulation Environment and Experimental Results

NS-2 simulator ver. 2.28 from [13] was used for the study of energy optimization of routing protocols like AODV and SAMPLE. The underlying MAC Protocol is defined by IEEE 802.11. Continuous bit rate (CBR) traffic sources along with TCP based traffic sources are used. The mobility model used was the Steady State Random Waypoint Mobility Model [20]. The field configurations are 1200 x 1200 m. The traffic generator script called cbrgen.tcl was used to generate CBR/TCP scenario of 8 sources at the rate of 4.0 kbps. The number of nodes in the simulation environment was 20 nodes for low density and 100 nodes for medium density. At least 5 scenarios files for Steady State Random Waypoint Model at different maximum speed of 5, 10, 15, 20, 25 sec were used for testing protocols like AODV and SAMPLE.

5.1 Energy Parameter Configuration

The energy model for energy consumption is based on the parameters given in Table 1.

Tx	Rx	Idle	Sleeping
1400mW	1000mW	830mW	130mW

Table 1: Energy Consumption Parameters.

The initial energy supplied to all wireless nodes were 1000 Joules.

5.2 Energy Conservation Metric

Average Energy Conservation (AEC) – This metric specifies the average of residual energy of the nodes in an ad-hoc network at the finish of the simulation period.

$$\sum_{i=1}^{N} E_i / N \qquad (7)$$

where E_i is the remaining energy and N is the number of nodes in the given ad-hoc network

5.3 Simulation Results

The results of the energy optimization framework which consists of SAMPLE integrated with SPAN on 802.11 and compared with SAMPLE and AODV on 802.11 are illustrated in Figure 5, Figure 6, Figure 7 and Figure 8.

In Figure 5 it is observed that the average energy conservation is the least for the adaptive reinforcement learning based routing algorithm SAMPLE. AODV with low density of nodes and on CBR based traffic outscores SAMPLE by 7% on an average. Thus to improve the energy conservation issues we apply an energy aware extension SPAN to aid modified SAMPLE. In Figure 5 it is clearly observed that SAMPLE-SPAN outperforms AODV and SAMPLE by 25 % and 30% respectively. The high link duration of RW-LD aids in link stability and possibly lesser routing overhead, which in turns aids energy conservation.

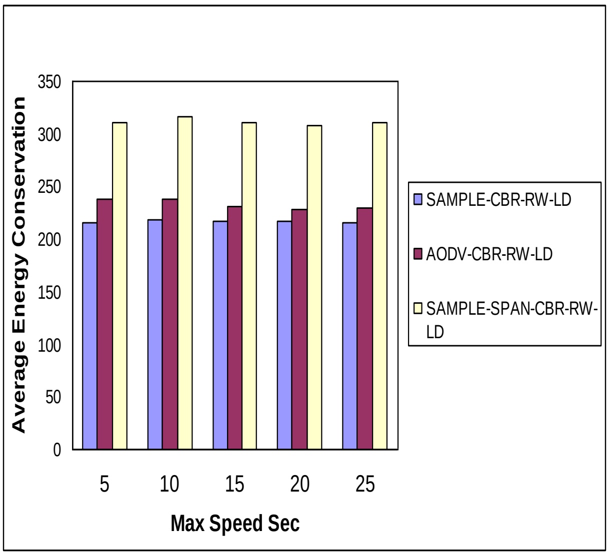

Figure 5: Average energy conservation of SAMPLE, AODV and SAMPLE-SPAN with CBR based traffic and on low-density of nodes.

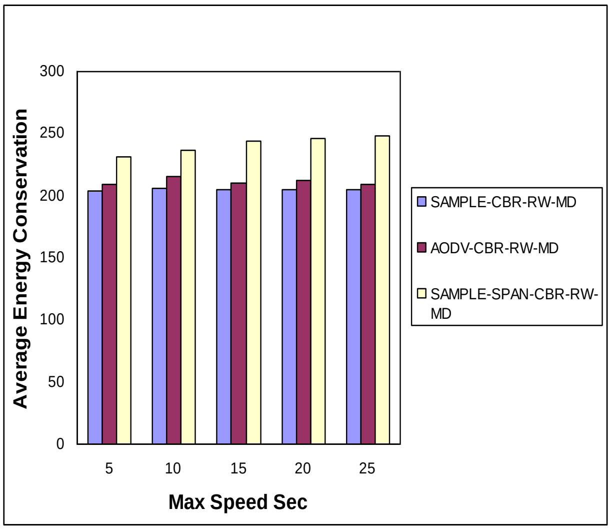

Figure 6: Average energy conservation of SAMPLE, AODV and SAMPLE-SPAN with CBR based traffic and on medium-density of nodes.

In Figure 6 it is seen that SAMPLE once again performs least in terms of energy conservation and is inferior as compared to AODV by 3% on an average. AODV and SAMPLE performance in terms of energy conservation is more or less the same because of improved spatial dependency exhibited by RW-MD, which reduces network partitions in moderately scalable setting and this in turn complements energy conservation. SAMPLE-SPAN outperforms AODV and SAMPLE on an average by 15% and 13% respectively.

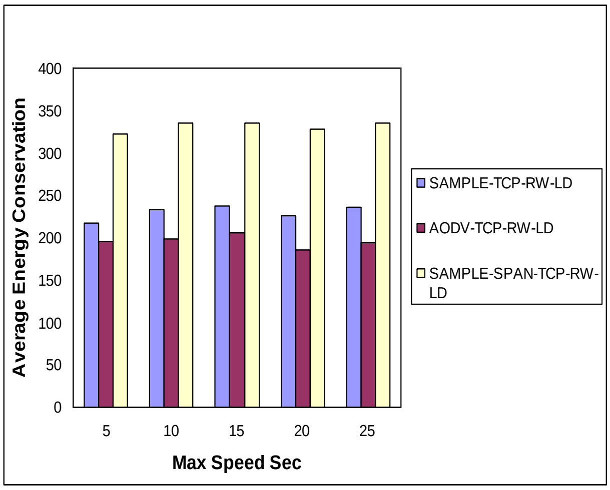

Figure 7: Average energy conservation of SAMPLE, AODV and SAMPLE-SPAN with TCP based traffic and on low-density of nodes.

In Figure 7 for TCP based traffic and low density of nodes, which have stable link properties, SAMPLE with its inherent capability to extract stable links and model them continuously outperforms AODV by 15%. SAMPE-SPAN further conserves energy and outperforms AODV and SAMPLE by 41% and 31% respectively

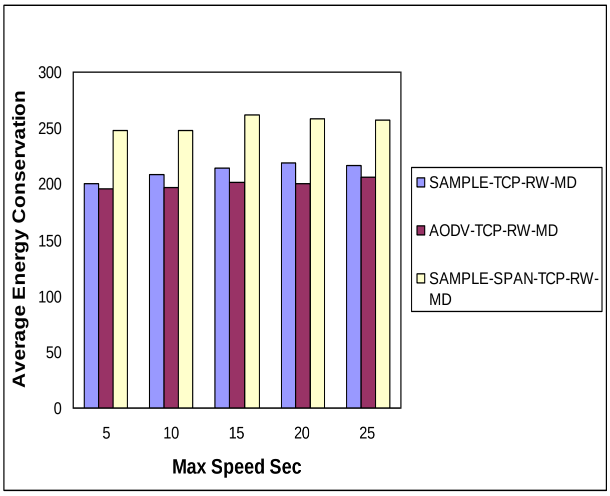

Figure 8: Average energy conservation of SAMPLE, AODV and SAMPLE-SPAN with TCP based traffic and on medium-density of nodes.

In Figure 8 under moderately scalable condition and with relatively higher degree of spatial dependency, it is observed that SAMPLE on TCP based traffic outperforms AODV by 6% and SAMPLE-SPAN outperforms AODV and SAMPLE by 22% and 17% respectively.

To prove the statistical significance of our results we conducted Hypothesis Testing on the simulation results. The assumptions are as follows. The null hypothesis H_0 states that the values for both the data set are similar i.e. $H_0 : \mu1 = \mu2$. The alternative hypothesis Ha states that the values for both the data set are not similar i.e. Ha : $\mu1 \neq \mu2$, whereby we would like to prove that there is sufficient energy conservation achieved by SAMPLE with SPAN extensions. The Hypothesis t-Test [24] results using Analysis Toolpak[23] are illustrated in Table 2, Table 3, Table 4 and Table 5.

	Variable 1	*Variable 2*
Mean	233.0138	311.2544
Variance	19.00113	7.840572
Observations	5	5
Pooled Variance	13.42085	
Hypothesized Mean Difference	0	
Df	8	
t Stat	-33.7685	
P(T<=t) one-tail	3.23E-10	
t Critical one-tail	1.859548	
P(T<=t) two-tail	6.46E-10	
t Critical two-tail	2.306004	

Table 2: Hypothesis t-Test: Two-Sample Assuming Equal Variances for comparison of average energy conservation of AODV and SAMPLE-SPAN for low density and CBR Traffic.

	Variable 1	*Variable 2*
Mean	211.1304	241.650 1
Variance	7.115177	50.2387 5
Observations	5	5
Pooled Variance	28.67696	
Hypothesized Mean Difference	0	
df	8	
t Stat	-9.01123	
P(T<=t) one-tail	9.18E-06	
t Critical one-tail	1.859548	
P(T<=t) two-tail	1.84E-05	
t Critical two-tail	2.306004	

Table 3 Hypothesis t-Test: Two-Sample Assuming Equal Variances for comparison of average energy conservation of AODV and SAMPLE-SPAN for Medium density and CBR Traffic.

	Variable 1	*Variable 2*
Mean	196.1009	331.38 28
Variance	53.8941	33.708 04
Observations	5	5
Pooled Variance	43.80107	
Hypothesized Mean Difference	0	
df	8	
t Stat	-32.3197	

P(T<=t) one-tail	4.58E-10	
t Critical one-tail	1.859548	
P(T<=t) two-tail	9.15E-10	
t Critical two-tail	2.306004	

Table 4: Hypothesis t-Test: Two-Sample Assuming Equal Variances for comparison of average energy conservation of AODV and SAMPLE-SPAN for Low density and TCP Traffic.

	Variable 1	*Variable 2*
Mean	199.9108	254.5423
Variance	17.32532	45.22516
Observations	5	5
Pooled Variance	31.27524	
Hypothesized Mean Difference	0	
df	8	
t Stat	-15.4459	
P(T<=t) one-tail	1.53E-07	
t Critical one-tail	1.859548	
P(T<=t) two-tail	3.07E-07	
t Critical two-tail	2.306004	

Table 5 Hypothesis t-Test: Two-Sample Assuming Equal Variances for comparison of average energy conservation of AODV and SAMPLE-SPAN for Medium density and TCP Traffic.

The test data was collected by varying mobility speed for 5, 10, 15, 20 and 25 m/sec. For each test the significance level was 0.05%. Our tests rejected the null hypothesis and this proved our simulation results with 95% confidence interval, for average energy conservation were significant for SAMPLE with SPAN extensions, as compared to average energy conservation for AODV routing protocol.

6 Conclusions

In this paper our study augmented and modified a reinforcement learning routing algorithm SAMPLE's energy conservation capabilities by integrating it with SPAN a cross layer energy saving extension in the proposed energy conservation optimization framework. We also analysed the experimental mobility model, the Steady State Random Waypoint Mobility Model for performance analysis under low and medium density of nodes to gain insight into energy conservation and scalability issues. From our experimental studies we observed that SAMPLE-SPAN outperformed AODV and SAMPLE on both low density and high density and for both CBR and TCP based traffic. Thus energy optimization should not be the property of only the network layer and it should be visualized at the data link layer also. We also verified our results statistically. In future we would like to apply the energy optimization frame work to an optimized mobility model to gain insights into superior mobility complementing energy conservation issues.

References

[1] S Corson and J Marker, Mobile Ad Hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations, RFC 2501, pp. 3-9, January 1999

[2] AEthem, Introduction to Machine Learning, Prentice Hall of India, pp. 370-375, 2006.

[3] S.A. Kulkarni, Dr. G.R Rao, "Modeling Reinforcement Learning Algorithms for Performance Analysis", *International Conference on Advances in Computing, Communication and Control - ICAC'03*, January 23-24th, 2009.

[4] J Dowling, E Curran, R Cunningham, and V Cahill, Using Feedback in Collaborative Reinforcement Learning to Adaptively Optimize MANET Routing, *IEEE Transactions on Systems Man and Cybernetics. Part A: Systems and Humans*, Vol. 35, No.3, 2005

[5] S.A Kulkarni and Dr. G.R Rao, "Optimization of Reinforcement Learning Based Routing Algorithm Applied to Dynamic Network", *In Proceedings of ICCVR-2010*, August 19th to21st, 2010.

[6] B Chen, K Jamieson, H Balakrishnan and R Morris, "Span: An Energy-Efficient Coordination Algorithm for Topology Maintenance in Ad Hoc Wireless Networks", *ACM/IEEE International Conference on Mobile Computing and Networking (Mobi Comm. 2001)*, Rome, Italy, July 16-21, 2001.

[7] Wireless LAN Medium Access Control and Physical Layer Specifications, *IEEE 802.11 Standard* (IEEE Computer Society LAN MAN standards Committee), August 1999.

[8] S P Chaudri, J-Y L Boudec, M Vojnovic. Perfect Simulations for Random Trip Mobility Models. *Proceedings of the 38th Annual Symposium on Simulations*, pp. 72-79, 2005

[9] R Kravets, C Sengul, "Energy Conservation", In Ad-Hoc Networks Technologies and Protocols (Eds.) P Mohapatra and S Krishnamurthy, Springer, pp. 153-155, 2005.

[10] J-Y L Boudec and M Vojnovi´c, "Perfect simulation and stationarity of a class of mobility models," *in Proceedings of IEEE Infocom,* 2005.

[11] F Bai, N Sadagopan and A Helmy, "The IMPORTANT framework for analyzing the Impact of Mobility on Performance of RouTing protocols for AdhocNeTworks", *Elsevier Ad Hoc Networks*, 1, pp. 383-403, 2003.

[12] S.A. Kulkarni, G R Rao, "Mobility and Traffic Model Analysis for Vehicular Ad-hoc Networks", In Advances in Vehicular Ad Hoc Networks: Developments and Challenges, Ed. Mohamed Watfa, IGI Global, pp. 214-232, 2010.

[13] The Network Simulator – NS-2, http://www.isi.edu/nsam/ns/

[14] P Nurmi. Reinforcement Learning for Routing in Ad Hoc Networks. *Proc. 5thIntl. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks* (WiOpt, Limassol, Cyprus, April 2007) IEEE Computer Society, 2007

[15] B Chen and C Hwa Chang, Mobility Impact on Energy Conservation of Ad Hoc Routing Protocols, *SSGR* 2003, Italy, July 28-August 2, 2003.
[16] V Naumov, T Gross, “Scalability of routing methods in ad hoc networks”, *In Performance Evaluation Journal (Elsevier Science),* Volume 62, pp. 193-207, 2005.
[17] Y Xu, J Heidemann and D Estrin, “Geography-informed Energy Conservation for Ad Hoc Routing. *In proceedings of the Seventh Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, Rome, Italy, July 2001.
[18] C Raghavendra and S Singh, “PAMAS: Power Aware Multi-Access Protocol with Signaling for Ad Hoc Networks”, *ACM Computer Communication Review*, pp. 5- 26, July, 1998.
[19] S Singh, M Woo and C S Raghavendra, “Power-Aware Routing in Mobile Ad Hoc Networks”, *In Proceedings of the Fifth Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, Dallas, 1998.
[20] S P Chaudhuri, ns-2 Code for Random Trip Mobility Model, http://monarch. s.rice.edu/~santa /research/mobility/
[21] C Perkins, E Royer and S Das, “Ad Hoc on demand distance vector (AODV) routing”, http://www.ietf.org/internet-drafts/draft-ietf-manet-aodv-03.txt
[22] C Perkins and E Royer, “Ad hoc on-demand Distance Vector Routing”, *In Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications*, pp. 90-100, Feb 1999
[23] Analysis Toolpak, http://office.microsoft.com /en-us/excel-help/about-statistical-analysis-tools-HP00 5203873.aspx
[24] K N Krishnaswamy, A I Shivakumar and M Mathirajan “Management Research Methodology – Integration of Principles, Methods and Techniques”, Pearson Education, pp. 352-355, 2006.
[25] S A Kulkarni, G R Rao, “A Performance Analysis of Energy Efficient Routing in Mobile Ad Hoc Network”, *International Journal of Simulations – Systems, Science and Technology – IJSSST*: Vol. 10, No.1-A. . pp. 1- 9, February – 2010.
[26] J Yoon, M Liu and B Noble, “Random Waypoint Considered Harmful”, *in Proceedings IEEE INFOCOM 2003,* San Francisco, C A, April 2003.
[27] S P Chaudhri, J-Y L Boudec, M Vojnovic, “Perfect Simulations for Random Trip Mobility Models”, *in Proceedings of the 38th Annual Symposium on Simulation*, 2005, pp. 72-79.

Simulation of Multiphase Thermo-Fluid Phenomena by a Local Meshless Numerical Approach

Gregor Kosec
Institute Jozef Stefan, Slovenia
E-mail: gkosec@ijs.si, http://comms.ijs.si/~gkosec/

Thesis Summary

Keywords: multiphase, fluid flow, heat transfer, meshless, local, parallel, OpenMP, pressure-velocity

Received: April 20, 2012

In this paper, the summary of the doctoral thesis focused on the local meshless based solution procedure for solving a multiphase thermo-fluid flow is presented.

Povzetek: V članku je predstavljen povzetek doktorskega dela, namenjenega reševanju večfaznega toplotno kapljevinskega toka preko lokalne brezmrežne numerične metode.

1 Introduction

The computational modelling of multiphase systems has become a highly popular research subject due to its pronounced influence in better understanding of nature as well as in the development of advanced technologies. Melting of the polar ice caps, the global oceans dynamics, various weather systems, water transport, soil erosion and denudation, magma transport and manufacturing of nano-materials, improving casting processes, energetic studies, exploitation of natural resources, welding, casting and advanced solidifications are typical contemporary examples where multiphase systems play an important role. The understanding and modelling of more and more complex physical systems allows the community to address important issues, like environmental consequences of some specific action, on one side and improving technological processes, on the other side. The understanding of natural processes is a key factor in improving our relation to the environment and the quality of life with better and cleaner industrial capacities. Several natural and technological phenomena fall into category of multiphase thermo-fluid flow problems and presented thesis addresses some of them.

In the dissertation [1] a new numerical approach towards solving multi-phase thermo-fluid problems is treated. The volume averaged governing equations for mass, energy, momentum, and species transfer on the macroscopic level, together with the species transfer on the microscopic level are considered [2]. The main issues in solving such physical models occur due to strong nonlinearities and the strong couplings. Such situations are dealt within the dissertation by solving a spectrum of benchmark problems.

2 Methodology

The involved systems of governing equations are solved by local meshless [3] technique based on the Local Radial Basis Function Collocation Method (LRBFCM) [4]. In the thesis, the classical LBRFCM is enhanced with several new functionalities. The basis functions selection and approximation type are generalized. The influence domain selection is dynamic and depends on the node distribution topology. The local approximation system is stabilized by shaping the basis functions. An effective algorithm for adding and/or removing discretization nodes on/from the computational domain is presented. The introduced features increase the applicability and stability of the original LRBFCM.

To minimize computational cost and maximize parallelization efficiency, basic idea throughout the dissertation is to keep the locality of the algorithm. The strong form and the explicit time scheme are used. Furthermore, a local approach to the fluid flow computation is required. Respectively, a local pressure-velocity coupling algorithm is introduced, which allows construction of completely local mass conservation corrections. The proposed pressure-velocity coupling is based on the mass continuity violation, where the pressure correction is proportionally linked to the divergence of the computed velocity field. Some of the cases in the dissertation deal with highly convective dominant problems that might behave unstable. An adaptive upwind strategy [5] is incorporated in the solution procedure in order to stabilize such situations.

The presented solution procedure is almost ideally parallelizable, as it is completely local. There is minimal communication with other parts of the computational domain. It is shown that for shared memory systems the OpenMP parallelization is trivial and effective [6].

The bulk part of the work needed to produce the results was implementation of the solution procedure in the C++ programming language.

3 Presentation of Results

To assess the proposed numerical method a spectra of tests are performed. The spatial discretization convergence, the time discretization convergence and the deformation of the regular node distribution impact on the accuracy are tested on two-dimensional diffusion equation with Dirichlet jump. Dirichlet jump problem has been also used to compare different numerical approaches (FEM, MLPG and FDM) with the presented LRBFCM.

The performance of the method is in the convection regimes assessed on two original tests, developed from the classical Smith and Hutton test [7]. The first test is focused on the impact of the node distribution density jump. The second test is dedicated to the comparison of the dynamic node distribution with the adaptive upwind strategy. Additional test of the nodes adaptivity approach is done by solving the two dimensional Burgers' equation [8] and comparing the results against published data.

The proposed local pressure-velocity coupling algorithm is tested on de Vahl Davis benchmark test [9] and tall cavity, natural convection in the porous media [10], double diffusive natural convection and melting of a pure material driven by a natural convection. To assess the characteristics of the proposed numerical approach numerous analyses and comparisons with the published data are performed. The comprehensive verification procedure shows good agreements with the previously published data, based on different numerical methods.

The final part of the dissertation represents the application of the proposed numerical model and developed solution procedure in a macrosegregation simulation [11]. The solidification of a binary alloy (Al-4.5%Cu) is considered. The results are verified against the classical FVM approach. The simulations of the macrosegregation are for the first time presented with perfect agreement of two entirely different numerical methods.

4 Conclusion

The represented achievements establish better understanding of the meshless numerical methods and broaden their application. The gained knowledge is expected to be directly applied in scientific and technological problems, where the multicomponent and multiphase fluid flow plays an important role. The developed code is written with a lot effort put in the optimization and the parallelization thus it can be used for treatment of more complex situations.

Acknowledgement

I would like to thank the Slovenian Research Agency for support in the framework of the projects Young Researcher Programme 1000-06-310232.

References

[1] Kosec, G.: *http://www.ung.si/~library/doktorati/fizika/10Kosec.pdf,* 2011

[2] Ni, J. & Beckermann, C., A volume-averaged two-phase model for transport phenomena during solidification. *Metallurgical and Materials Transactions B,* 22B, pp. 349-361, 1991.

[3] Liu, G.R., *Mesh Free Methods*, CRC Press: Boca Raton, 2003.

[4] Šarler, B., From global to local radial basis function collocation method for transport phenomena. *Advances in Meshfree Techniques,* pp. 257-282, 2007.

[5] Lin, H. & Atluri, S. N., Meshless local Petrov Galerkin method (MLPG) for convection-diffusion problems. *CMES: Computer Modeling in Engineering & Sciences,* 1, pp. 45-60, 2000.

[6] Kosec, G., Trobec, R., Depolli, M. & Rashkovska, A.: Multicore parallelization of a meshless PDE solver with OpenMP, ParNum 11, 2011.

[7] Smith, R. M. & Hutton, A. G. , The numerical treatment of advection: A performance comparison of current methods. *Numerical Heat Transfer,* A5, pp. 439-461, 1982.

[8] Burgers, J. M., A Mathematical Model Illustrating the Theory of Turbulence. *Advances in Applied Mechanics,* Volume 1, pp. 171-199, 1948.

[9] de Vahl Davis, G., Natural convection of air in a square cavity: a bench mark numerical solution. *International Journal of Numerical Methods in Fluids,* 3, pp. 249-264, 1983.

[10] Bertrand, O., Binet, B., Combeau, H., Couturier, S., Delannoy, Y., Gobin, D., Lacroix, M., Le Quéré, P., Médale, M., Mencinger, J., Sadat, H. & Vieira, G., Melting driven by natural convection. *International Journal of Thermal Sciences,* 38, pp. 5-26, 1998.

[11] Kosec, G., Založnik, M., Šarler, B. & Combeau, H., A Meshless Approach Towards Solution of Macrosegregation Phenomena. *Computers, materials & continua,* 580, pp. 1-27, 2011.

CALL FOR PAPERS

Information Society 2012

15th International Multiconference

8–12 October 2012, Ljubljana, Slovenia

IS 2012

http://is.ijs.si

The concepts of information society, information era, infosphere and infostress have by now been widely accepted. But what does it really mean for the society, science, technology, education, governments, our lives? The Information Society multiconference deals with information technologies, which are of major importance for the development of Europe. IS 2012 will serve as a forum for the world-wide and national community to explore further research directions, business opportunities and governmental policies. For these reasons we host a scientific meeting in the form of a multiconference, which will consist of several independent conferences with themes essential for the development of the information society. The main objective is the exchange of ideas and developing visions for the future of information society. IS 2012 is a high-quality multidisciplinary conference covering major recent scientific achievements.

Independent conferences within IS 2012

- Intelligent Systems
- 100 Years of Alan Turing and 20 Years of SLAIS
- Cognitive Sciences
- Collaboration, Software and Services in Information Society
- Data Mining and Data Warehouses (SiKDD 2012)
- Education in Information Society
- Facing Demographic Challenges
- FORSEE – Technological Forecasting in ICT
- Language Technologies
- Robotics

Submission and dates

Submitted papers must be original and not currently under review by another conference or journal. They should address the topic of one of the conferences within IS 2012, and should preferably be accessible to a wide audience. They can be in English or in Slovene, depending on the conference, and should be up to four pages long in most cases. All papers will be peer-reviewed. Both printed and electronic proceedings on a USB flash drive will be available.

Deadlines for paper submission:

- 30 July 2012 – Collaboration, Software and Services in Information Society
- 10 August 2012 – Language Technologies
- 31 August 2012 – Intelligent Systems, 100 Years of Alan Turing and 20 Years of SLAIS, Data Mining and Data Warehouses
- 1 September 2012 – Robotics
- 5 September 2012 – Facing Demographic Challenges

Organization

Main organizer: Jožef Stefan Institute, Department of Intelligent Systems

Organization committee: Matjaž Gams (chair), Mitja Luštrek (co-chair), Vedrana Vidulin, Vesna Koricki Špetič, Mitja Lasič and Robert Blatnik

JOŽEF STEFAN INSTITUTE

Jožef Stefan (1835-1893) was one of the most prominent physicists of the 19th century. Born to Slovene parents, he obtained his Ph.D. at Vienna University, where he was later Director of the Physics Institute, Vice-President of the Vienna Academy of Sciences and a member of several scientific institutions in Europe. Stefan explored many areas in hydrodynamics, optics, acoustics, electricity, magnetism and the kinetic theory of gases. Among other things, he originated the law that the total radiation from a black body is proportional to the 4th power of its absolute temperature, known as the Stefan–Boltzmann law.

The Jožef Stefan Institute (JSI) is the leading independent scientific research institution in Slovenia, covering a broad spectrum of fundamental and applied research in the fields of physics, chemistry and biochemistry, electronics and information science, nuclear science technology, energy research and environmental science.

The Jožef Stefan Institute (JSI) is a research organisation for pure and applied research in the natural sciences and technology. Both are closely interconnected in research departments composed of different task teams. Emphasis in basic research is given to the development and education of young scientists, while applied research and development serve for the transfer of advanced knowledge, contributing to the development of the national economy and society in general.

At present the Institute, with a total of about 900 staff, has 700 researchers, about 250 of whom are postgraduates, around 500 of whom have doctorates (Ph.D.), and around 200 of whom have permanent professorships or temporary teaching assignments at the Universities.

In view of its activities and status, the JSI plays the role of a national institute, complementing the role of the universities and bridging the gap between basic science and applications.

Research at the JSI includes the following major fields: physics; chemistry; electronics, informatics and computer sciences; biochemistry; ecology; reactor technology; applied mathematics. Most of the activities are more or less closely connected to information sciences, in particular computer sciences, artificial intelligence, language and speech technologies, computer-aided design, computer architectures, biocybernetics and robotics, computer automation and control, professional electronics, digital communications and networks, and applied mathematics.

The Institute is located in Ljubljana, the capital of the independent state of **Slove**nia (or S♡nia). The capital today is considered a crossroad between East, West and Mediterranean Europe, offering excellent productive capabilities and solid business opportunities, with strong international connections. Ljubljana is connected to important centers such as Prague, Budapest, Vienna, Zagreb, Milan, Rome, Monaco, Nice, Bern and Munich, all within a radius of 600 km.

From the Jožef Stefan Institute, the Technology park "Ljubljana" has been proposed as part of the national strategy for technological development to foster synergies between research and industry, to promote joint ventures between university bodies, research institutes and innovative industry, to act as an incubator for high-tech initiatives and to accelerate the development cycle of innovative products.

Part of the Institute was reorganized into several high-tech units supported by and connected within the Technology park at the Jožef Stefan Institute, established as the beginning of a regional Technology park "Ljubljana". The project was developed at a particularly historical moment, characterized by the process of state reorganisation, privatisation and private initiative. The national Technology Park is a shareholding company hosting an independent venture-capital institution.

The promoters and operational entities of the project are the Republic of Slovenia, Ministry of Higher Education, Science and Technology and the Jožef Stefan Institute. The framework of the operation also includes the University of Ljubljana, the National Institute of Chemistry, the Institute for Electronics and Vacuum Technology and the Institute for Materials and Construction Research among others. In addition, the project is supported by the Ministry of the Economy, the National Chamber of Economy and the City of Ljubljana.

Jožef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
Tel.:+386 1 4773 900, Fax.:+386 1 251 93 85
WWW: http://www.ijs.si
E-mail: matjaz.gams@ijs.si
Public relations: Polona Strnad

INFORMATICA

AN INTERNATIONAL JOURNAL OF COMPUTING AND INFORMATICS

INVITATION, COOPERATION

Submissions and Refereeing

Please submit a manuscript at: http://www.informatica.si/Editors/PaperUpload.asp. At least two referees outside the author's country will examine it, and they are invited to make as many remarks as possible from typing errors to global philosophical disagreements. The chosen editor will send the author the obtained reviews. If the paper is accepted, the editor will also send an email to the managing editor. The executive board will inform the author that the paper has been accepted, and the author will send the paper to the managing editor. The paper will be published within one year of receipt of email with the text in Informatica MS Word format or Informatica LaTeX format and figures in .eps format. Style and examples of papers can be obtained from http://www.informatica.si. Opinions, news, calls for conferences, calls for papers, etc. should be sent directly to the managing editor.

QUESTIONNAIRE

☐ Send Informatica free of charge

☐ Yes, we subscribe

Please, complete the order form and send it to Dr. Drago Torkar, Informatica, Institut Jožef Stefan, Jamova 39, 1000 Ljubljana, Slovenia. E-mail: drago.torkar@ijs.si

Since 1977, Informatica has been a major Slovenian scientific journal of computing and informatics, including telecommunications, automation and other related areas. In its 16th year (more than eightteen years ago) it became truly international, although it still remains connected to Central Europe. The basic aim of Informatica is to impose intellectual values (science, engineering) in a distributed organisation.

Informatica is a journal primarily covering intelligent systems in the European computer science, informatics and cognitive community; scientific and educational as well as technical, commercial and industrial. Its basic aim is to enhance communications between different European structures on the basis of equal rights and international refereeing. It publishes scientific papers accepted by at least two referees outside the author's country. In addition, it contains information about conferences, opinions, critical examinations of existing publications and news. Finally, major practical achievements and innovations in the computer and information industry are presented through commercial publications as well as through independent evaluations.

Editing and refereeing are distributed. Each editor can conduct the refereeing process by appointing two new referees or referees from the Board of Referees or Editorial Board. Referees should not be from the author's country. If new referees are appointed, their names will appear in the Refereeing Board.

Informatica is free of charge for major scientific, educational and governmental institutions. Others should subscribe (see the last page of Informatica).

ORDER FORM – INFORMATICA

Name: ....

Title and Profession (optional): ....

....

Home Address and Telephone (optional): ....

....

Office Address and Telephone (optional): ....

....

E-mail Address (optional): ....

Signature and Date: ....

Informatica WWW:

http://www.informatica.si/

Referees from 2008 on:

Informatica

An International Journal of Computing and Informatics

Web edition of Informatica may be accessed at: http://www.informatica.si.

Subscription Information Informatica (ISSN 0350-5596) is published four times a year in Spring, Summer, Autumn, and Winter (4 issues per year) by the Slovene Society Informatika, Vožarski pot 12, 1000 Ljubljana, Slovenia.
The subscription rate for 2012 (Volume 36) is
– 60 EUR for institutions,
– 30 EUR for individuals, and
– 15 EUR for students
Claims for missing issues will be honored free of charge within six months after the publication date of the issue.

Typesetting: Borut Žnidar.
Printing: ABO grafika d.o.o., Ob železnici 16, 1000 Ljubljana.

Orders may be placed by email (drago.torkar@ijs.si), telephone (+386 1 477 3900) or fax (+386 1 251 93 85). The payment should be made to our bank account no.: 02083-0013014662 at NLB d.d., 1520 Ljubljana, Trg republike 2, Slovenija, IBAN no.: SI56020830013014662, SWIFT Code: LJBASI2X.

Informatica is published by Slovene Society Informatika (president Niko Schlamberger) in cooperation with the following societies (and contact persons):
Robotics Society of Slovenia (Jadran Lenarčič)
Slovene Society for Pattern Recognition (Franjo Pernuš)
Slovenian Artificial Intelligence Society; Cognitive Science Society (Matjaž Gams)
Slovenian Society of Mathematicians, Physicists and Astronomers (Bojan Mohar)
Automatic Control Society of Slovenia (Borut Zupančič)
Slovenian Association of Technical and Natural Sciences / Engineering Academy of Slovenia (Igor Grabec)
ACM Slovenia (Dunja Mladenič)

Informatica is surveyed by: ACM Digital Library, Citeseer, COBISS, Compendex, Computer & Information Systems Abstracts, Computer Database, Computer Science Index, Current Mathematical Publications, DBLP Computer Science Bibliography, Directory of Open Access Journals, InfoTrac OneFile, Inspec, Linguistic and Language Behaviour Abstracts, Mathematical Reviews, MatSciNet, MatSci on SilverPlatter, Scopus, Zentralblatt Math

The issuing of the Informatica journal is financially supported by the Ministry of Higher Education, Science and Technology, Trg OF 13, 1000 Ljubljana, Slovenia.

Volume 36 Number 2 June 2012 **ISSN 0350-5596**

Informatica

An International Journal of Computing and Informatics

Informatica **36** (2012) Number 2, pp. 119–225