Conference of PhD Students in Computer Science

Volume of extended abstracts

CS^2

Organized by the Institute of Informatics of the University of Szeged

June 29 –July 2, 2010
Szeged, Hungary

Scientific Committee:

János Csirik (Chair, SZTE)
Lajos Rónyai (Co-chair, SZTAKI)
András Benczúr (ELTE)
Hasszan Charaf (BME)
Tibor Csendes (SZTE)
László Cser (BCE)
János Demetrovics (ELTE)
József Dombi (SZTE)
Zoltán Ésik (SZTE)
Ferenc Fiedler (PE)
Zoltán Fülöp (SZTE)
Aurél Galántai (ÓE)
Tibor Gyimóthy (SZTE)
Zoltán Horváth (ELTE)
Csanád Imreh (SZTE)
Zoltán Kató (SZTE)
Zoltán Kása (Sapientia EMTE)
László Keviczky (SZIE)
János Kormos (DE)
László Kozma (ELTE)
János Levendovszky (BME)
Eörs Máté (SZTE)
Attila Pethő (DE)
András Recski (BME)
Tamás Roska (PPKE)
Endre Selényi (BME)
Tamás Szirányi (SZTAKI)
Tibor Tóth (ME)

Organizing Committee:

Kálmán Palágyi, Balázs Bánhelyi, Tamás Gergely, Zoltán Kincses

Address of the Organizing Committee

c/o. Kálmán Palágyi
University of Szeged, Institute of Informatics
H-6701 Szeged, P.O. Box 652, Hungary
Phone: +36 62 546 396, Fax: +36 62 546 397
E-mail: cscs@inf.u-szeged.hu
URL: http://www.inf.u-szeged.hu/~cscs/

Sponsors

Novadat Bt., Győr
Polygon Publisher
TÁMOP-4.2.3 program of the Hungarian National Development Agency
University of Szeged, Institute of Informatics

Preface

This conference is the seventh in a series. The organizers have tried to get together those PhD students who work on any fields of computer science and its applications to help them possibly in writing their first abstract and paper, and may be to give their first scientific talk. As far as we know, this is one of the few such conferences. The aims of the scientific meeting were determined on the council meeting of the Hungarian PhD Schools in Informatics: it should

- provide a forum for PhD students in computer science to discuss their ideas and research results,
- give a possibility to have constructive criticism before they present the results in professional conferences,
- promote the publication of their results in the form of fully refereed journal articles, and finally
- promote hopefully fruitful research collaboration among the participants.

The best talks will be awarded with the help of our sponsors. The papers emerging from the presented talks will be forwarded to the journals of Acta Cybernetica (Szeged), and Periodica Polytechnica (Budapest) and the mathematics oriented papers to Publicationes Mathematicae (Debrecen). The deadline for the submission of the papers is the end of August 2010. The manuscripts will be forwarded to the proper journals.

Although we did not advertise it on the web, a high number of good quality abstracts have been submitted. If you encounter any problems during the meeting, please do not hesitate to contact one of the Organizing Committee members. The organizers hope that the conference will be a valuable contribution to the research of the participants, and wish a pleasant stay in Szeged.

Szeged, June 2010

Kálmán Palágyi
Balázs Bánhelyi
Tamás Gergely
Zoltán Kincses

Contents

Program

Tuesday, June 29

- 08:00 - 08:45 Registration
- 08:45 - 09:00 Opening
- 09:00 - 10:40 Talks - Software Engineering 1 (5x20 minutes)
- 10:40 - 11:00 Break
- 11:00 - 11:50 Plenary talk
- 11:50 - 12:00 Break
- 12:00 - 12:40 Talks - Web (2x20 minutes)
- 12:40 - 14:20 Lunch at the Gödör restaurant
- 14:20 - 16:00 Talks - Image and Singal Processing 1 (5x20 minutes)
- 16:00 - 16:20 Break
- 16:20 - 18:00 Talks - Applications 1 (5x20 minutes)
- 18:00 Dinner at the Alabárdos restaurant

Wdnesday, June 30

- 09:00 - 10:40 Talks - Networks and Fuzzy (5x20 minutes)
- 10:40 - 11:00 Break
- 11:00 - 11:50 Plenary talk
- 11:50 - 12:00 Break
- 12:00 - 12:40 Talks - Databases and Stochastic Processes (2x20 minutes)
- 12:40 - 14:20 Lunch at the Gödör restaurant
- 14:20 - 16:00 Talks - Image and Signal Processing 2 (5x20 minutes)
- 16:00 - 16:20 Break
- 16:20 - 18:00 Talks - Optimization 1 (5x20 minutes)
- 18:00 Reception at the 1st floor

Thursday, July 1

- 09:00 - 10:40 Talks - Matematics of Computing (5x20 minutes)
- 10:40 - 11:00 Break
- 11:00 - 11:50 Plenary talk
- 11:50 - 12:00 Break
- 12:00 - 12:40 Talks - Numerical Ananlysis (2x20 minutes)
- 12:40 - 14:20 Lunch at the Gödör restaurant
- 14:20 - 15:20 Talks - Optimization 2 (3x20 minutes)
- 15:20 - 15:40 Break
- 15:40 - 16:40 Talks - Distributed Computing (3x20 minutes)
- 16:40 - 17:40 Free program
- 17:40 Social program (visit the Informatic Collection and gala dinner)

Friday, July 2

- 09:00 - 10:40 Talks - Software Engineering 2 (5x20 minutes)
- 10:40 - 11:00 Break
- 11:00 - 11:50 Plenary talk
- 11:50 - 12:00 Break
- 12:00 - 12:40 Talks - Applications 2 (2x20 minutes)
- 12:40 - 14:20 Lunch at the Gödör restaurant
- 14:20 - 15:40 Talks - Computer Graphics (4x20 minutes)
- 15:40 Closing

Detailed Program

Tuesday, June 29

8:00	**Registration**
8:45	**Opening**
Section	**Software Engineering 1**
9:00	Gy.I. Szabó: *How Much is XML Involved in DB Publishing?*
9:20	Z. Szűgyi: *DeepTest the Static Debugger of Java Programs*
9:40	G. Tóth: *Comparison of Programmers' Opinion in Change Impact Analysis*
10:00	F. Fischer: *The Comparison of Hybrid Impact Sets Computed During the Application in Different Areas*
10:20	D. Tengeri: *Introducing Database Slicing Method and Context-Free Export/Import*
10:40	Break
Section	**Plenary talk**
11:00	B. Buchberger: *Can Inventions in Mathematics and Computer Science be Automated?*
11:50	Break
Section	**Web**
12:00	R. Hodován: *Speeding up browsing with caching regular expressions*
12:20	V. Bilicki, M. Kasza, M. Jelasity: *Small Degree BitTorrent*
12:40	Lunch at the Gödör restaurant
Section	**Image and Singal Processing 1**
14:20	L. Varga, P. Balázs, A. Nagy: *Object Rotation Effects on Binary Tomographic Reconstruction*
14:40	N. Hantos, P. Balázs: *Median Filtering in Algebraic Reconstruction Methods*
15:00	B. Antal: *Approaches to Improve Macula Detection in Retinal Images*
15:20	K. Bolla: *Distinguish Competitor Robot Swarms with On-board Visual Perception System*
15:40	I. Lázár: *Feature Selection and Classification Algorithms for Retinal Microaneurysm Detection*
16:00	Break
Section	**Applications 1**
16:20	B. Torma, L. Gerencsér: *Economic Interpretation of GARCH Models: an Agent-Based Simulation Study*
16:40	K. Veress: *Parameter Estimation of a Flow-Measurement Function Used in Digital Angiography*
17:00	A. Huszti, P. Pleva: *An Ideal Anonymous E-exam System*
17:20	J. Bozsik: *Decision Tree Combined with Neural Networks for Financial Forecast*
17:40	G. Bényász, L. Cser: *Clustering Financial Time Series on CUDA*
18:00	Dinner at the Alabárdos restaurant

Wednesday, June 30

Section	**Networks and Fuzzy**
9:00	B. Hullár, S. Laki, A. György, G. Vattay: *New Methods in Payload Based Network Traffic Classification*
9:20	N. Sterbinszky: *Benchmarking the Relationship of Database Management Systems and Application Servers on the Web*
9:40	A. Bóta, M. Krész, A. Pluhár: *Applications of Overlapping Community Detection*
10:00	E. Griechisch: *Comparison of Clustering and Community Detection Algorithms*
10:20	S.K. Manouchehr: *Active Control of Cable Bridges Using Fuzzy Logic*
10:40	Break
Section	**Plenary talk**
11:00	A.E. Eiben: *Evolutionary Algorithms and What to Do About Their Parameters*
11:50	Break
Section	**Databases and Stochastic Processes**
12:00	R. Giachetta, Zs. Máriás: *Performance Evaluation of Storing Inhomogeneous Descriptive Data of Digital Maps*
12:20	J. Hunor: *Guided Exploration in Policy Gradient Algorithms with Gaussian Process Function Approximation*
12:40	Lunch at the Gödör restaurant
Section	**Image and Signal Processing 2**
14:20	T.S. Tasi, P. Balázs: *Obtaining Geometrical Properties of Binary Images from two Projections Using Neural Networks*
14:40	M. Gara, P. Balázs: *Binary Tomographic Reconstruction with an Object-Based Evolutionary Algorithm*
15:00	G. Németh, P. Kardos: *Algorithms for Topological Kernel Extraction and Thinning*
15:20	P. Kardos: *Order-Independent Sequential Thinning on Two-Dimensional Binary Images*
15:40	G. Páli, Z. Gilián, Z. Horváth, A. Persson: *Describing Digital Signal Processing Systems in A Functional Reactive Style*
16:00	Break
Section	**Optimization 1**
16:20	A. Mosavi: *The Multiobjective Optimization Package of IOSO; Applications and Future Trends*
16:40	E. Palatinus: *Circle Covering and its Application for Telecommunication Networks*
17:00	L. Pál, T. Csendes: *Traffic Flow Prediction in Service Networks*
17:20	A. Csernenszky, Gy. Kovács, M. Krész, A. Pluhár: *Parameter Optimization of Infection Models*
17:40	T. Bartók, Cs. Imreh: *Vehicle Routing with Multi-Dimensional Loading Constraints*
18:00	Reception at the 1st floor

Thursday, July 1

Section	**Matematics of Computing**
9:00	P. Varga: *Characterization of Semi-CNS Polynomials*
9:20	Zs. Lengyel, S. Sike: *Analysis of ODF Document Schema Transformations*
9:40	T. Demián: *Common Logic*
10:00	M. Hegyháti: *Mathematical Models and Methodologies in Batch Process Scheduling*
10:20	K. Kalauz, B. Bertók, F. Friedler: *Algorithmic Synthesis of Structural Alternatives for Business Processes*
10:40	Break
Section	**Plenary talk**
11:00	A. Galántai: *Reliability of Numerical Algorithms*
11:50	Break
Section	**Numerical Ananlysis**
12:00	A. Gáti: *Miller Analyser for Matlab — A Matlab Package for Automatic Roundoff Analysis*
12:20	I. Bársony: *Surprises with Different C/C++ Implementations and a Quizzical MAPLE Example (Two Short Case Studies in Programming)*
12:40	Lunch at the Gödör restaurant
Section	**Optimization 2**
14:20	V. Árgilán, J. Balogh, J. Békési, B. Dávid, M. Krész, A. Tóth: *Heuristic Approaches for Scheduling Problems in Public Transportation*
14:40	A. Bekéné Rácz: *Preprocessing in Linear Fractional Programming*
15:00	L. Pál, T. Csendes: *Interval Based Sensor Network Localization*
15:20	Break
Section	**Distributed Computing**
15:40	R.O. Legéndi, Attila Szabó: *Distributing Agent-Based Simulation*
16:00	A. Szabó, R.O. Legéndi: *An Agent-Based Model Formalization to Support Distributed Execution*
16:20	V. Vad: *Resampling of Volumetric Data with GPU Generated Distance Field*
16:40	Free program
17:40	Social program (visit the Informatic Collection and gala dinner)

Friday, July 2

Section	**Software Engineering 2**
9:00	Z. Ujhelyi: *Static Type Checking of Model Transformation Programs*
9:20	Á. Csendes: *Survey of Dynamic Voltage Scaling Methods for Energy Efficient Embedded Systems*
9:40	K. Bunyik, K. Markó: *The Vector Library Testing of Feldspar Programming Language Based on Axiomatic Semantics*
10:00	M. Tóth, I. Bozó: *Building Dependency Graph for Slicing Erlang Programs*
10:20	I. Szombath: *Online Infrastructure Dependency Detection and Tracking*
10:40	Break
Section	**Plenary talk**
11:00	M. Kovács: *An Introduction to Fractional Calculus*
11:50	Break
Section	**Applications 2**
12:00	K. Veress: *Radiocommunication Testbed for Wireless Sensor Networks*
12:20	T. Dobján, Sz. Pletl, T. Deák, L. Doszpod, G. Pór: *Identification of the Place and Materials of Knocking Objects in Flow Induced Vibration*
12:40	Lunch at the Gödör restaurant
Section	**Computer Graphics**
14:20	G. Klár, G. Valasek: *Employing Pythagorean Hodograph Curves for Artistic Patterns*
14:40	L. Lócsi: *Calculating Non-Equidistant Discretizations Generated by Blaschke Products*
15:00	R. Kunkli, M. Hoffmann: *Skinning of Circles Using Biarcs*
15:20	R. Szendrei, I. Elek, I. Fekete, M. Márton: *A Knowledge-Based Approach to Raster-Vector Conversion of Large Scale Topographic Maps*
15:40	Closing

Approaches to Improve Macula Detection in Retinal Images

Bálint Antal

Diabetic retinopathy (DR) is the most common cause of blindness in the developed countries. The screening of this disease is currently done manually, nevertheless it is a very resource demanding procedure. Several efforts have been made to establish a computer-based automatic screening system. The basis of an automatic screening system is the analysis of retinal images.

The analysis consists of two parts: the detection of certain anatomical parts and the detection of disorders caused by DR. In this paper, we present an approach which belongs to the first category. Macula is the center of sharp vision, so disorders which appear within this anatomical part can lead to vision loss. That is, the proper detection of the macula is essential also for a DR screening system to be able to classify lesions closer to the macula center as more dangerous.

In this paper, we present a novel macula detector, which is competitive with the state-of-the-art methods (e. g. [1]). This approach is based on the fact that macula appears as a spot which is darker than the surface of the retina. The algorithm can be summarized as follows:

First, we extract the green plane from the color fundus image. We generate the background image by applying median filter and subtract it from the green plane, resulting in a shade corrected image. Next, we binarize the image by considering all non-zero pixels as foreground pixels, and others as background. Finally, we apply an image labeling procedure and select the largest component as the macula.

In addition, we present the effect of determining the parameter setup for the algorithm automatically using a simulated annealing algorithm. We also determine the spatial bias of the algorithm, the correction of which leads to further improvement. We also show that this algorithm improves the accuracy of an ensemble-based system[2] which uses several macula detectors.

Acknowledgments

This work was supported in part by the János Bolyai grant of the Hungarian Academy of Sciences, and by the TECH08-2 project DRSCREEN - Developing a computer based image processing system for diabetic retinopathy screening of the National Office for Research and Technology of Hungary (contract no.: OM-00194/2008, OM-00195/2008, OM-00196/2008). We also acknowledge the Moorefields Eye Hospital, London for their clinical support.

References

[1] S. Sekhar, W. Al-Nuaimy, A. K. Nandi, Automated localization of optic disc and fovea in retinal fundus images, 16th European Signal Processing Conference, Lausanne, Switzerland, 5 pages, 2008.

[2] B. Harangi, R.J. Qureshi, A. Csutak, T. Peto, A. Hajdu: Automatic detection of the optic disc using majority voting in a collection of optic disc detectors, 7th IEEE International Symposium on Biomedical Imaging (ISBI 2010), Rotterdam, The Netherlands, 2010, 1329-1332.

Heuristic Approaches for Scheduling Problems in Public Transportation

Viktor Árgilán, János Balogh, József Békési, Balázs Dávid, Miklós Krész, and Attila Tóth

Scheduling problems arising in public transportation are complex tasks. Considering both theoretical and practical aspects, the planning process of a vehicle scheduling system consists of four main sub-problems: vehicle scheduling, vehicle assignment, crew scheduling, and crew rostering. These can be solved as separate problems. Many studies have dealt with them, and a number of solution algorithms and methods have been published [2, 3, 4]. Though these methods give good local solutions for the sub-problems respectively, we need to solve them in a given order to attain the global solution of the scheduling problem. This does not guarantee a global solution, moreover in extreme cases – because of the fact that the methods do not consider the constraints of other sub-problems - no feasible solution exists at all. This leads to the introduction of combined approaches (see eg. [5]), but considering the fact that all sub-problems are NP-hard, these can only be used on smaller instances. Modeling such a system is a really hard and complex problem even for a middle sized-city (like in our benchmark case, Szeged, Hungary: 160.000 inhabitants, the problem having 2763 trips, 107 vehicles and 162 drivers for a regular workday). However, good solution can be given using a proper heuristic approach.

Our approach considers vehicles and drivers together. For this, a vehicle schedule is needed (given by the time-space network [6]), and based on this, a vehicle assignment has to be introduced [1]. The most crucial point of the crew scheduling is the rule of special specified breaks. The method transforms the schedules given by the vehicle assignment in such a way, that the rules regarding the breaks can be applied. According to their length, the schedules are divided into classes that are based on the maximum number of working hours assigned to a driver. The schedules can be broken down into sub-schedules, which can be inserted together to form new schedules, and trips can be moved from a schedule to another as well. If these operations still don't result in a new schedule where the breaks can inserted, well-chosen trips have to be removed to get a schedule which satisfies the rules for the breaks. The trips removed are appended to a list, and the whole process has to be started over again from the vehicle scheduling step, using the members of the list as an input, until there are no trips left on the list.

The results so far are promising, the method leads to a decrease both in the number of applied vehicles and employed drivers as opposed to the present practice used by the local bus company of Szeged.

Acknowledgements

This work was partially supported by Szeged City Bus Company (Tisza Volán, Urban Transport Division).

References

[1] Balogh, J., Békési, J., Galambos, G., Krész, M.: An Assignment Model for Real-World Vehicle Scheduling Problem with Refueling, in Proceedings of the 9th Workshop on Models and Algorithms for Planning and Scheduling Problems, Abbey Rolduc, The Netherlands, June 29 - July 03, 2009, pp. 229

[2] Bodin, L., Golden, B., Assad, A., Ball, M.: Routing and Scheduling of Vehicles and Crews: The State of the Art. Computers and Operations Research, 10, 63-212., 1983

[3] Daduna, J.R., Paixao, J.M.P.: Vehicle scheduling for public mass transit – an overview. In: J.R. Daduna et al., Editors, Computer-aided transit scheduling, LNEMS vol. 430, Springer, Berlin (1995), pp. 76–90.

[4] A.T. Ernst, H. Jiang, M. Krishnamoorthy and D. Sier, Staff scheduling and rostering: A review of applications, methods and models, European Journal of Operational Research 153, 3–27, 2004

[5] Huisman, D. , Freling, R., Wagelmans, A.P.M., Multiple-Depot Integrated Vehicle and Crew Scheduling, Transportation Science, v.39 n.4, p.491-502, 2005

[6] Kliewer, N., Mellouli, T., Suhl, L.: A time-space network based exact optimization model for multi-depot bus shceduling. European Journal of Operational Research, 175, 1616-1627, 2006

Surprises with Different C/C++ Implementations and a Quizzical MAPLE Example (Two Short Case Studies in Programming)

István Bársony

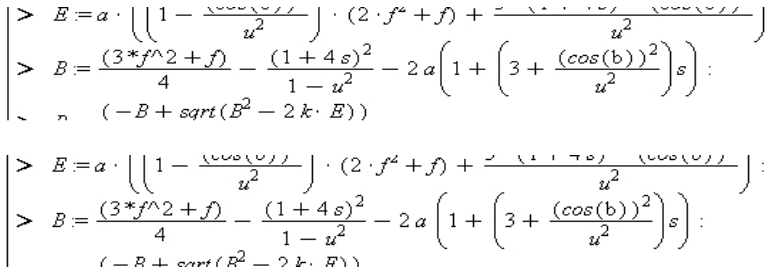
The C/C++ language has different implementations (Microsoft Visual C++, GNU C++, BorlandC++, Dev-C++,...). The C/C++ language is known as a creditable, well defined one. So the output of the program detail below can be a little bit unexpected:

```
void kiiro(int a, int b, int c, int d, int e, int f);
void main(){
int i=3;
kiiro(i,++i,i++,++i,++i,i);
}
void kiiro(int a, int b, int c, int d, int e, int f)
{printf("%d %d %d %d %d %d\n",a,b,c,d,e,f);}
```

When we compile and run it with different C/C++ implementations, the outputs will be different. As first case study we examine this phenomenon at assembly level using the development environment provided by vendors. The differences are well traceable on obtained ASSEMBLY sources. Moreover trying to calculate the next or some similar expression

```
a=i + i++ + ++i + i++;
```

we can get new sudden results. As a second case we present an effect in MAPLE. Why give these lines (see below the line with $B :=$) syntactically good but semantically quite different results (that is, the result is wrong in one of them)?



We have got the reason of this strange behavior in a very special way.

Concrete results: Different C/C++ compilers can give different result for the same source code.

References

[1] Geddes K., Labahn G., Monagan M.:Maple 12 Advanced Programming Guide, Maplesoft 2008, ISBN:978-1-897310-47-2

[2] Josuttis N. M.: The C++ Standard Library: A Tutorial and Reference, Addison-Wesley Professional; 1 edition (August 22, 1999), ISBN-13: 978-0201379266

Vehicle Routing with Multi-Dimensional Loading Constraints

Tamás Bartók and Csanád Imreh

One of the most widely researched problem of operation research is the vehicle routing problem (VRP), which also has numerous real life applications and many mathematical variations. There are also many exact solving and heuristic algorithms available in the literature for each of these variations. The majority of these models do not consider the loading of the vehicle, and those few that do consider, have mostly only a one-dimensional loading constraint, like weight, volume or fragility constraints and these models do not care with the real geometric loading of the items along the given route. According to our knowledge, the usage of models in practice, that take the geometric loading and other constraints into account at the same time, is very limited, mostly because of its high combinatorial complexity.

In this talk we present a vehicle routing model, which is a combined optimization of the geometric loading freight into vehicles, while not exceeding the weight constraint and the routing of the vehicles along a road network, with the aim of serving customers with minimum travelling cost. We investigate, how such loading and routing is possible for a given set of 3-dimensional items and a given set of vehicles.

Since this problem is a generalization of two well-known, NP-hard problems, the capacitated vehicle routing problem (CVRP) and the three-dimensional bin packing problem (3DBPP), this problem is also NP-hard. Because of this reason, we can not expect to provide a fast (polinomial) algorithm, that could solve this problem to optimality in a reasonable time for larger, real-world problem intstances. Therefore, we provide a heuristic solution, that does not provide an optimal solution, but an acceptable one in a reasonable time. We'll also provide measurements for our solution with empirical analysis.

Preprocessing in Linear Fractional Programming

Anett Bekéné Rácz

Preprocessing is very important for solving large optimization problems irrespectively of using interior point or simplex algorithm. Most of the professionally developed solvers automatically use preprocessing techniques to maintain numerical stability and improve performance. Even though computers have become even faster, the real-world models have increased in size. The reason can be the complexity and the model generators too. The aim of preprocessing is to reduce the problem size and to find redundancy and the unbounded or infeasible problems. In this paper we describe the main results of our investigations connected with preprocessing techniques in Linear-Fractional Programming (LFP). Our investigations are based on the use of well-known preprocessing techniques of linear programming and we adapt them to LFP Problems. Some of these techniques can be used in LFP without any changes, but the others have to be adapted. Sometimes this adaptation is not so obvious. Not only the preprocessing but also the postsolve techniques are different in nonlinear environment. The paper presents some preprocessing techniques with its postsolve operations based on [1], [2] and gives an overview its adaptation into LFP.

References

[1] Cs. Mészaros, U. H. Suhl. Advanced preprocessing techniques for linear and quadratic programming, *OR Spectrum vol. 25, p575-595. 2003.*

[2] E. D. Andersen, K. D. Andersen. Presolving in linear programming, *Mathematical Programming vol. 71, p221-245. 1995.*

Clustering Financial Time Series on CUDA

Gábor Bényász and László Cser

Clustering in financial time-series databases has received significant attention lately. Differently from the normal clustering algorithm the main challenge of the time-series clustering is the high dimensionality. The fact, which helped to overcome this awkwardness, was the development of the time series models which enabled the utilization of clustering of time series by compressing the dimensionality of time series into parameterised expression which can be compared and clustered.

The first models based upon moving average and autoregression of the time series called AR, MA and ARIMA [2] models. There were introduced some clustering model [1], [4] those ARIMA time-series which are intend to use the Linear Predictive Coding (LPC) cepstrum of time series as Euclidean distance between the LPC cepstrals of the time series. More developed model of time-series, called GARCH [3], models take also account the autoregressive conditional heteroscedasticity of the time series and therefore is able to interpret nonstationarity time-series. As clustering of GARCH models of time series has not been investigated so far the next logical step could have been to cluster the GARCH based time-series, even DWT (Discrete Wavelet Transformation) was chosen for modelling time-series which can be clustered after that. We chosed that because of the shortcomings of the GARCH models as it cannot handle adequately and entirely the shock-like movements of the time-series (which is essential character of the financial data) which followa long-tail distribution. Similarly DFT (Discrete Furier Transformation) has been also heavily investigated [6] [7] for time series clustering, but [8] showed that DWT is significantly faster to model and provide a multi-resolution decomposition.

DWT has been intensively used in technical and natural sciences for decades, but using it for financial purposes is a very new initiative. Using DWT time series are projected into the time-frequency plane of sliding windows. The Wavelet coefficients hold the compressed characterization of the time series and form the dimension of time-series clustering. The major differences of using the DWT for comparing the time-series are rooted in the usage of the coefficients (first k, last k, largest k, adaptive [9] coefficients) for further reducing the dimensionality because in high dimensional spaces the distance between the nearest and the farthest neighbour gets increasingly smaller [5], making it impossible and meaninglessly to cluster. We consider the use the adaptive method discussed in [9] to reduce the number of the coefficients. As a side effect the high dimensionality of the clustering algorithm can slow down so that even becomes useless. To overcome this performance related difficulties CUDA was intend to use as clustering algorithms could be massively parallelised [10, 11, 12, 13]. Not knowing the initials numbers of the clusters we considered using iterative k-Means algorithm. To enhance the implementation the wavelet transformation was also carried out via CUDA as Wavelet algorithms are naturally parallel. For example, if enough processing elements exist, the wavelet transform for a particular spectrum can be calculated in one step by assigning a processor for every two points. The proposed algotrithm are attempted to run and evaluate using of closing prices of stocks listed on the NYSE. The method for the determination of the sliding window is still subject of the further investigation.

References

[1] Konstantinos Kalpakis, Dhiral Gada, and Vasundhara Puttagunta: Distance Measures for Effective Clustering of ARIMA Time-Series. *In the Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM'01), San Jose, CA, November 29-December 2, 2001, pp. 273-280.*

[2] Brockwell, Peter J. and Davis, Richard A.: *Time Series: Theory and Methods, 1987, Springer-Verlag.*

[3] Engle R. F.: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom ination. *Econometrica 50, 1987, 987-1007.*

[4] Bagnall A. J., Janacek G. J.: Clustering time series from ARMA models with clipped data. *In Proceedings of the Tenth ACM SIGKDD international Conference on Knowledge Discovery and Data Mining (Seattle, WA, USA, August 22 - 25, 2004). KDD '04. ACM, New York, NY, 49-58.*

[5] Beyer K., Goldstein J, Ramakrishnan K, Uri S.: When is nearest neighbor meaningful? *Lecture Notes in Computer Science, 1540:217-235, 1999.*

[6] Agrawal R, Faloutsos C., and Swami A. N.: Effcient Similarity Search In Sequence Databases. *In D. Lomet, editor, Proceedings of the 4th International Conference of Foundations of Data Organization and Algorithms (FODO), pages 69-84, Chicago, Illinois, 1993. Springer Verlag.*

[7] Rafiei D. and Mendelzon O. A.: Similarity-based queries for time series data. *In Proceedings of the 1997 ACM SIGMOD international Conference on Management of Data (Tucson, Arizona, United States, May 11 - 15, 1997). J. M. Peckman, S. Ram, and M. Franklin, Eds. SIGMOD '97. ACM, New York, NY, 13-25.*

[8] Wu Y, Agrawal D.,Abbadi A. E.: A comparison of DFT and DWT based similarity search in time-series databases. *In CIKM, pages 488-495, 2000.*

[9] Morche F.: Time series feature extraction for data mining using DWT and DFT, *whitepaper, 2003*

[10] YING, L.: High performance computing with CUDA, *Conference of High Performance Computing, XVII, 2009)*

[11] Zechner M, Granitzer M.: Accelerating K-Means on the Graphics Processor via CUDA, *pp.7-15, 2009 First International Conference on Intensive Applications and Services, 2009*

[12] Shuai, C. et al: Performance Study of General Purpose Appliaction on GPU using Cuda, *Journal of Distributed and Paralell computing, XVI, 2009.*

[13] Shalom, S. A., Dash, M., and Tue, M.: Efficient K-Means Clustering Using Accelerated Graphics Processors. *In Proceedings of the 10th international Conference on Data Warehousing and Knowledge Discovery (Turin, Italy, September 02 - 05, 2008). I. Song, J. Eder, and T. M. Nguyen, Eds. Lecture Notes In Computer Science, vol. 5182. Springer-Verlag, Berlin, Heidelberg, 166-175*

Small Degree BitTorrent

Vilmos Bilicki, Miklós Kasza, and Márk Jelasity

It is well-known that the BitTorrent file sharing protocol is responsible for a significant portion of the Internet traffic. A large amount of work has been devoted to reducing the footprint of the protocol in terms of the amount of traffic [1], however, its flow level footprint has not been studied in depth. We argue in this paper that the large amount of flows that a BitTorrent client maintains will not scale over a certain point. To solve this problem, we first examine the flow structure through realistic simulations. We find that only a few TCP connections are used frequently for data transfer, while most of the connections are used mostly for signaling. This makes it possible to separate the data and signaling paths. We propose that, as the signaling traffic provides little overhead, it should be transferred on a separate dedicated small degree overlay while the data traffic should utilize temporal TCP sockets active only during the data transfer. Through simulation we show that this separation has no significant effect on the performance of the BitTorrent protocol while we can drastically reduce the number of actual flows.

Keywords: peer-to-peer; BitTorrent; small degree overlay; TCP

Acknowledgements

M. Jelasity was supported by the Bolyai Scholarship of the Hungarian Academy of Sciences.

References

[1] M. Slot, P. Costa, G. Pierre and V. Rai: Zero-Day Reconciliation of BitTorrent Users with Their ISPs, *Euro-Par 2009 Parallel Processing*, 2009.

Distinguish Competitor Robot Swarms with On-board Visual Perception System

Kálmán Bolla

A robot swarm consists of a number of physical robots, where each robot as autonomous system works together, communicates with other team members. One of the major tasks in a robot swarm is the team member detection in various environments meanwhile doing its tasks and sharing information with other swarm members. A possible solution to this problem is provided now. Our kin recognition system is based on that the members of the swarm have an on-board visual perception system and each robot is masqueraded with a movement invariant zebra pattern. The visual kin recognition method [1] employs Fourier analysis of this zebra pattern which has got a reliable working in various environment. The Fourier transform is able to emphasize the features of zebra pattern and provides a good solution to pattern recognition. For the sake of simplicity we used the fast adaptation of Fourier transform. This method executes feature enhancement and feature selection on one captured image in complexity of $c{\cdot}n{\cdot}\log n$, where $n{\cdot}\log n$ is the complexity of the fast Fourier transform and c is the column sampling constant. A much more complex problem is when more competitor robot swarms are in the same environment. In this case kin recognition [1] is completed with a competitor robot recognition, too. In this abstract we assume that two competitor robot swarms are working in the same environment and the robots of the two swarms know about distinguishing marks of the other swarm. Distinguishing marks in this scope means two different zebra patterns, which have an effect on the different Fourier spectrum results. To distinguish competitor robot swarms and to resolve the problem two different feature selection functions are needed. These functions are actually threshold functions, which select the columns of pattern from the digital image and can identify the robot swarm objects. The threshold functions are determined empirically. These functions uniquely determine the swarm of a certain robot and distunguish it from the objects of the environment. Another scope of this method is the distance evaluation from the detected frequency. The byproduct of detection is a frequency value which grows along with the awaying zebra pattern. With the help of this method the robot-robot distances could be determined by the detected frequencies. Since the two different robot swarms generate different frequencies two distance measurements are needed for the distance evaluation. The result of the measurements shows a linear relationship between frequency and distance values thus the results can be approximated well by two linear functions. This method also provides a useful tool to distinguish competitor robot swarms including a distance evaluation feature.

References

[1] K. Bolla, T. Kovács, G. Fazekas. Compact Image Processing-based Kin Recognition, Distance Measurement and Identification Method in a Robot Swarm, *ICCC-CONTI 2010 Conference*, submitted.

Applications of Overlapping Community Detection

András Bóta, Miklós Krész, and András Pluhár

The analysis of networks is considered to be a traditional research field, which in the recent years has received a new impulse, thanks to the variety of available test databases and benchmarks. These graphs are so large in some cases, that only the fastest, near linear time algorithms have a chance of tackling with the given tasks. One of the popular approaches to network analysis, community detection has received a lot of attention in the recent years both from the point of theory and practical applications.

On the other hand, the concept of community doesn't have a strict definition. Thinking intuitively, one can imagine communities as subsets of nodes in a graph, in which the edge density is higher than between these sets. The traditional approach to community detection deals with disjoint communities. There exists, however another approach to communities: overlapping communities. In this case, the nodes in the graph can belong to multiple communities. Articles dealing with overlapping communities are scarce in the literature, and approach the subject mostly from the theoretical point of view. Just like traditional communities, overlapping communities also lack a strict definition. In most applications the overlaps between the communities can only happen at the edges of the given communities, another approach however allows greater overlaps, identifying a rich community structure in the original network. This can be an important source of information in itself, but the community graph built from the results can also form the basis of further analysis.

In most cases the data to be analyzed is in the form of unweighted graphs, but sometimes it is necessary to deal with the weighted counterparts, which can contain more information. Algorithms able to effectively utilize this additional information are also scarce both in literature and applications.

In this article we propose an improved version of an unpublished algorithm [1], which can detect overlapping communities in a weighted graph, and scales in quasi-linear time, thus allowing the analysis of large test databases. During this process the algorithm explores a dense community structure, which we use as the basis of further algorithms [2], like clustering, label propagation [3] and influence propagation [4]. We compare the results of this procedure with results from running these methods on the original, unmodified graph. Our test databases mostly come from the field of molecular biology, but we have tested our methods on graphs from different fields of science.

References

[1] L. Csizmadia: Community detection in social graphs, Master's Thesis, Szeged, 2003. (In Hungarian: Közösségfelismerés ismeretségi gráfokban)

[2] A. Pluhár: Fast algorithms for identifying communities in social graphs, Study, Szeged, 2002. (In Hungarian: Ismeretségi gráfok közösségeinek meghatározása gyors algoritmusokkal)

[3] U. Raghavan, R. Albert, S. Kumura: Near linear time algorithm to detect community structures in large-scale networks, Physical Review E 76, 2007.

[4] D. Kempe, J. Kleinberg,É. Tardos: Influential nodes in a diffusion model for social networks, Proceedings of ICALP 2005.

Decision Tree Combined with Neural Networks for Financial Forecast

József Bozsik

In this article I would like to introduce a hybrid adaptive method. There is high variety of methods for financial forecasts. This method is focusing for the financial default forecast in economics, but this method can be used generally for other financial forecasts as well, for example calculating value at risk.

This hybrid method combines two classical adaptive methods, which are the decision trees and the artificial neural networks. Many international articles were published all over the world about combining the two methods, but the speciality of this method which I would like to introduce is that this method uses the special issues of default forecasting in economics. The special knowledge is used as complementary heuristics in the hybrid model. The additional knowledge does not belong to the integral part of the model that is why the built model can be adapted easily for solutions of other problems. Of course in this case it is useful to build the given problem's speciality into the model. It is required from all well-chosen and well-sophisticated heuristics that the problem should be solved in an optimal way. This means that either higher accuracy and/or faster running time are expected.

In the article I would like to introduce the basics of decision trees and neural networks. These well-known structures can be definitely well applied in case of classification and forecast problems. In this article I will show a new model which is made by the combination of this two methods. In the beginning I will shortly introduce the well known perceptron and the multi layer perceptron model of neural networks and the ID3 algorithm which is used by decision trees. I will show the special points which are needed for the combination and I will show the classical combination form of these methods. In this part I would like to give a short outlook for the published articles from this area. After this part I would like to introduce in details the used hybrid method. I will show the new method's abstract model and the problems which occurred during building of the new model and the solution for it, for example I will show the case of over-teaching and the treatment of the continuous valued attributes. In order to solve the problems I used on the one hand the classical methods which can be found in the literature, on the other hand I established own methods using the problem's specialities and with this the problem could be solved.

The new $neural - tree$ method was built by using and testing of real company data. The results were compared to the results of a well-known and world-wide used economical default forecasting model. This method is the discriminance analysis. During the introduction of the results I would like to shortly summarize the basics of the discriminance analysis. For the testing it is necessary to introduce the discriminance analysis model. Without introducing all the details I will show the main steps of the model. The comparison of the models is made by using data from 2009.

I will show the results of the models in details in table form. Every measurement is done more times and by leaving out of the measurements the best and the worst results, I calculated the average of residual values which are published in this article. Analysing the results I will show the classification accuracy of the new method.

At the end of the article I will summarize the model's advantages and disadvantages and the barriers of the model. I will show the reasons of the classification accuracy of the results and explain the barriers. I will show the development possibilities of the model and the further research opportunities and questions in a schematic way.

References

[1] T.-S. Lim, W.-Y. Loh, and Y.-S. Shih, A comparison of prediction accuracy, complexity, and training time of 33 old and new classification algorithms, *Machine Learning*, vol. 40, pp. 203-228, 2000.

[2] S. Behnke and B. Karayiannis, Competitive neural trees for pattern classification, *IEEE Trans. Neural Networks*, vol. 9, pp. 1352-1369, 1998.

[3] O. T. Yildiz and E. Alpaydin, Linear discriminant trees, *17th Int. Conf. Machine Learning*, P. Langley, Ed. San Mateo, CA: Morgan Kaufmann, 2000, pp. 1175-1182.

The Vector Library Testing of Feldspar Programming Language Based on Axiomatic Semantics

Karina Bunyik and Krisztián Markó

This paper is a part of a project which defines a high level domain specific (DSL) language (Feldspar) embedded in Haskell that allows description of digital signal processing (DSP) algorithms and has special characteristics [2]. Vector library is the front-end of Feldspar, it contains standard list functions. Examples written in Feldspar are being compiled to a intermediate representation called core output. The advantage of this step is that the core language can be easily compiled to an imperative language [1]. The goal of the research was to test the Vector library of the Feldspar programming language and to check the equivalence of the Feldspar compiler and the Feldspar interpreter. The result of this paper is a testing method, which is checking the equality of the Feldspar interpreter and compiler with property based unit testing. The method is using a developed testing software which generates random data using QuickCheck and after that it verifies the Feldspar axioms [3]. Furthermore, we developed 80 axioms implemented in Feldspar, which describe the properties of the functions from the Vector library. Finally we ran each axiom for 100 random input data with the testing software, measured and interpreted the output. The research was done using the Glasgow Haskell Compiler (GHC), the Haskell QuickCheck property based testing tool and the Feldspar programming language, with its compiler and interpreter.

Acknowledgements

This work was supported by KMOP-2008-1.1.2, Ericsson Software Research and Ericsson Business Unit Networks SSF.

References

[1] D. Gergely, T. Mate, G. Zoltán, P. Gabor, N. Gyula, H. Zoltán, A. Emil, S. Mary, V. András, L. Bo and P. Anders. Efficient Code Generation from the High-level Domain-specific Language Feldspar for DSPs, *ODES: Workshop on Optimizations for DSP and Embedded Systems*, pages 12–20, 2010.

[2] A. Emil, C. Koen, D. Gergely, H. Zoltán, K. Karin Keijzer, L. Bo, P. Anders, S. Mary, S. Josef and V. András. Feldspar: A domain specific language for digital signal processing algorithms, *Memocode*, submitted, 2010.

[3] Koen Claessen and John Hughes. Quickcheck: A lightweight tool for random testing of Haskell programs, *ACM SIGPLAN Notices*, pages 268–279, 2000.

Survey of Dynamic Voltage Scaling Methods for Energy Efficient Embedded Systems

Áron Csendes

The power consumption of embedded devices is becoming more and more important, thus the energy efficiency needs to be optimized. Today's embedded hardware components (CPU, memory etc) make it possible to scale both their voltage level and their frequency dynamically in order to achive optimum energy consumption and meet computation time limitations at the same time.

In this paper we have collected some algorithms that use dynamic voltage and frequency scaling and have set comparison criteria to compare them. The comparison of the methods is done using XEEMU: an improved XScale power simulator.

References

[1] K. Choi, R. Soma, M. Pedram. Dynamic Voltage and Frequency Scaling based on Workload Decomposition, *ISLPED'04, August 9-11, 2004, Newport Beach, California, USA*

[2] G. Dhiman, T. S. Rosing. Dynamic Power Management Using Machine Learning, *ICCAD'06, November 5-9, 2006, San Jose, CA*

[3] Z. Herczeg, Á. Kiss, D. Schmidt, N. Wehn, T. Gyimóthy. XEEMU: An Improved XScale Power Simulator, *PATMOS 2007, LNCS 4644, pp. 300-309, 2007*

[4] P. Pillai, K. G. Shin. Real-Time Dynamic Voltage Scaling for Low-Power Embedded Operating Systems, *SOSP'01, October 21-24, 2001, Chateau Lake Louise, Banff, Canada*

Parameter Optimization of Infection Models

András Csernenszky, Gyula Kovács, Miklós Krész, and András Pluhár

Recently one of the main directions of datamining is the study and use of network data. Our research is concentrated on the network data about the links of big and medium size of companies that can be deduced from the bank transactions. A year ago the main goal for us was to develop an universal model to predict churn and bankruptcy. That work resulted in the implementation of the Domingos-Richardson cascade model for bankruptcy forecasting in 2009 September, which has since performed in the Bank quite well. The aim of our current research is to re-examine and redefine parameter optimization options and handle this issue in a more elaborate way. The performance of the method greatly depends on the estimation of the infection probabilities. In this research we get the better influence values assigned to the edges (which are representing the connections between two companies) that futher improves the infection predictions. The cornerstones for these improvements are the involvement on certain static variables and an appropriate optimization on the parameters that take their effects into account.

References

[1] P. Domingos, M. Richardson, Mining the Network Value of Costumers.*7th Intl. Conf. on Knowl-edge Discovery and Data Mining*, 2001.

Common Logic

Tamás Demián

Common Logic (CL) is a framework for a family of logic languages, based on first-order logic with the purpose of standardizing abstract syntax and semantics for knowledge representation (KR) and information interchange. CL is very expressive because it has been designed as a superset of many different notations.

The framework permits and encourages the development of a variety of different syntactic forms, called dialects sharing a single semantics. A dialect may use any desired syntax, but it must be possible to demonstrate precisely how the concrete syntax of a dialect conforms to the abstract syntax and semantics, which are based on a model theoretic interpretation. CL specification does not include any concrete syntactic forms. Specific form left to the KR designers. Once syntactic conformance is established, a dialect gets the CL semantics for free, as they are specified relative to the abstract syntax only, and hence are inherited by any conformant dialect. That is why we can perform meaning-preserving translations between dialects. However, some CL dialects may be more expressive than others. A given dialect need not use all the features of CL.

Common Logic has some novel features, chief among them being a syntax which is signature-free, type-free if needed and permits 'higher-order' constructions such as quantification over classes or relations while preserving a first-order model theory, and a semantics which allows theories to describe intensional entities such as classes or properties. CL was designed for easy, natural use on the web, so it also includes numerous web-oriented features. CL does not specify the inference rules but provide tools for expressing them in the dialects.

In 2007 CL became an ISO standard publicly available at [2]. The standard also includes three sample dialects:

- CLIF - Common Logic Interchange Format, based on KIF
- CGIF - Conceptual Graph Interchange Format
- XCL - eXtended Common Logic Markup Language, based on XML

Other important languages could also be defined as dialects. Among them are the RDF and OWL languages (defined by the W3C), SQL, Prolog, OCL, Datalog, RuleML, Controlled English, Controlled Chinese, Controlled Spanish and UML.

In this talk we will discuss the abstract syntax, the semantics and the key features of CL. We will focus on the metamodel [1] of the abstract syntax using UML notation. We will also provide the concrete syntax (dialect) of Petri nets demonstrating the use of the CL approach.

References

[1] M. Ferenczi, A. Pataricza, and L. Rónyai, editors. Formal methods in computing. Akadémiai Kiadó, 2005.

[2] Information technology - Common Logic (CL): a framework for a family of logic-based languages ISO/IEC Standard, Reference number: 24707:2007(E). standards.iso.org/ittf/PubliclyAvailableStandards/c039175_ISO_IEC_24707_2007(E).zip

Identification of the Place and Materials of Knocking Objects in Flow Induced Vibration

Tibor Dobján, Szilveszter Pletl, Tamás Deák, László Doszpod, and Gábor Pór

Many industrial systems contain pipes with fluid flow in it, either for transmitting materials or for cooling purposes. If some parts of the system detached or loosened it may go to either chaotic or deterministic motion due to forces gained from the flow energy. This motion is called flow induced vibration. If loose part knocks on the inner surface of the tube (or other compartment) then audible knock is generated. These are surface waves on the metal surfaces. The place of the knocks and the knocking material are crucial from the point of view of the fate of the given industrial objects. Therefore identification of the event, finding its place and identification of knocking material have primary importance form the point of view of the safety and maintenance of the system.

We investigated the improvement of the identification of the event recognition using autoregressive modeling based filtering [1] and sequential probability ratio test. While some parts of this technique had been elaborated earlier, the realization of those algorithms in LABVIEW and its integration into an embedded system are brand new developments.

For the localization of the place of the event, we investigated basically the time delay estimation methods. Time delays were estimated using cross-correlation technique and impulse response estimation. From the estimated time delays using linear measures of the piping system one can conclude the place of the knocks. However, besides the direct effect, i.e. besides the shortest route of the sound there are several other routes, reflections in the system, which in the case of repeated knocks can overlap with the arriving new front of the subsequent knocks. Selection of the arriving front were based on the physical nature of different fronts, however, this required an intelligent programming technique.

To give a hint on the material of the knocking objects we estimate the auto power spectral density function by dividing the frequency band into high frequency part and low frequency part. It was shown, that the ratio of these partial RMS values are different for knocking object of different materials. We present the first results of division of the APSD into four parts. It can be clearly seen, that this may improve the identification.

Finally, we discuss the possibility of introduction of neural network techniques into the process of identification of the place and material of knocking objects.

References

[1] George E. P. BOX, Gwilym M. Jenkins: TIME SERIES ANALISIS forecasting and control, Holden-Day, 1976

The Comparison of Hybrid Impact Sets Computed During the Application in Different Areas

Ferenc Fischer

The impact analysis is an important topic in software engineering, due to the continuous changing of software. The impact set [1] is a dependency set which contains the impacting and/or the impacted program elements for a set of program elements. The impact set of changing can be used in several parts of the software lifecycle, for example to estimate the costs of software changing or to help programmers to find program artifacts, - which may have been affected by the changing - or to seek for occurred bugs (debugging). These tasks demand various impact sets in terms of size, safeness and cost. We defined and calculated various impact sets (hybrid impact sets) by combination and limitation of the static and the dynamic impact sets. Static impact sets are calculated by the analysis of the source code without executing of the program. Dynamic impact sets are dependencies occurring during program runs on inputs.

To determine the dynamic impact set we created execution trace and computed DFC (Dynamic Function Coupling [2]) relation between methods. During calculating the static impact set we statically analyzed the source code and built a representation of the code to retrieve SEASEB (union of Static Execute After and Static Execute Before [3]) relation.

We define different strategies to compute these impact sets based on the distance between methods. This distance can be used for changing the size, the safeness and the costs of combined impact sets by distance limitation and/or by subtype limitation (call, return, sequential relation). Based on Reliability Growth Modelling we can reduce the computation cost of the dynamic impact set by running only selected test cases.

We have computed these impact sets for several Java programs to characterize these sets. A selection strategy is given by us to choose the most suitable method and the most appropriate impact set for the necessary tasks, goals and for the available resources. This comparison of the hybrid impact sets based on DFC and SEASEB is unique and the selection strategy can help the programmers to choose the best hybrid impact set for their tasks and resources.

References

[1] V. Rajlich and P. Gosavi. Incremental change in objectoriented programming, *IEEE Software*, 21(4):62-69, 2004.

[2] Árpád Beszédes and Tamás Gergely and Szabolcs Faragó and Tibor Gyimóthy and Ferenc Fischer. The Dynamic Function Coupling Metric and Its Use in Software Evolution, *Proceedings of 11th European Conference on Software Maintenance and Reengineering (CSMR'2007)*, pages 103-112, 2007.

[3] Jász, Judit and Beszédes, Árpád and Gyimóthy, Tibor and Rajlich, Václav. Static Execute After/Before as a Replacement of Traditional Software Dependencies, *Proceedings of the 2008 IEEE International Conference on Software Maintenance (ICSM'08)*, pages 137-146, 2008.

Binary Tomographic Reconstruction with an Object-Based Evolutionary Algorithm

Mihály Gara and Péter Balázs

The aim of *Binary Tomography (BT)* is to determine the interior of homogeneous objects from their projections, without destroying or damaging them. In BT usually only few projections are available – in our work we used only four angles to take the projection data. Owing to the limited number of projections some a priori information is also necessary to ensure accurate reconstruction. In our work we suppose that the two-dimensional cross-sections of the studied objects consist of disjoint disks inside a ring. This special feature inspires us to represent the image using an object-based model. Each image is represented by a list containing the center and the radius of the disks and the outer ring. With such representation the reconstruction can be performed with an evolutionary algorithm.

Evolutionary Algorithms have an initial population and they usually use two kinds of operators, *Mutation* and *Crossover*. With their aid the further generations are created. Mutation can modify a disk by moving or resizing it or, it also can increase or decrease the number of disks in the image. Crossover interchanges the disk-list-segments of the two images. We work with a fixed number of entities in all generations (we evaluate them with the so-called fitness-function and only keep the bests). The fitness function is based on the projection values, objects having projections similar to the ones of the original object have better fitness value.

To improve the algorithm the fitness function can be modified to hold additional prior information such as the number of disks constituting the image. To predict this number we used the well-known C4.5 *Decision Tree* classifier and then, the fitness of the entities in the population was biased according to the (uncertain) classification.

In this paper we present the details of the above-mentioned reconstruction method and our experimental results. This paper is based on our previous work [1].

References

[1] P. Balázs and M. Gara. An Evolutionary Approach for Object-Based Image Reconstruction Using Learnt Priors, *Lecture Notes in Comput. Sci.*, 5575:520–529, 2009.

Miller Analyser for Matlab — A Matlab Package for Automatic Roundoff Analysis

Attila Gáti

Miller Analyser for MATLAB is an automatic roundoff error analyser software, that extends the work of Miller et al. [1, 2, 3]. The software runs within the MATLAB environment, and can test the stability of numerical methods given as m-functions. Based on the algorithm of Miller, a number $\omega(d)$ is associated with each set d of input data. The function $\omega(d)$ measures rounding error, i.e. $\omega(d)$ is large exactly when the method applied to d produces results which are excessively sensitive to rounding errors. A numerical maximizer is applied to search for large values of ω. Finding large values of ω can be interpreted, that the given numerical method is suffering from a specific kind of instability.

We can perform analysis based on several error measuring numbers (various ways of assigning ω), and beside analysing the propagation of rounding errors in a single algorithm we can also compare the numerical stability of two competing numerical methods, which neglecting rounding errors compute the same values.

The analysis is based on the standard model of floating point arithmetic, which assumes that the individual relative rounding errors on arithmetic operations are bounded by the machine rounding unit. Practically, the computed result equals the correctly rounded exact result. The IEEE 754/1985 standard of floating point arithmetic guarantees that the standard model holds for addition, subtraction, multiplication, division and square root. Unfortunately, it is not true for the exponential, trigonometrical, hyperbolical functions and their inverses. Hence, we can analyse only numerical algorithms that can be decomposed to the above mentioned five basic operations and unary minus, which is considered error-free.

The first step of computing the error function $\omega(d)$ is building the computational graph of the analysed numerical method. Decomposition of a numerical method at a particular input $d = d_0$ to the allowed arithmetic operations give rise to a directed acyclic graph, the computational graph, with a node for each input value, output value and operation. There are arcs from each arithmetic node (ie., one corresponding to an operation) to the nodes for its operands and from each output node to the operation that computes its value.

According to the resulting computational graph the output computed as a function: $R_{d_0}(d, \delta)$ ($R_{d_0} : \mathbb{R}^{n+m} \to \mathbb{R}^k$), where $d \in \mathbb{R}^n$ is the input vector, and δ is the vector of individual relative rounding errors on the m arithmetic operations ($\delta \in \mathbb{R}^m$, $\|\delta\|_\infty \leq u$, where u is the machine rounding unit). The computation of $\omega(d)$ is based on the partial derivatives of R_{d_0} with respect to the input and the rounding errors. We apply automatic differentiation on the graph in reverse order, ie. the chain rule is applied in the opposite direction as the basic operations were executed.

References

[1] Webb Miller, *Software for roundoff analysis*, ACM Trans. Math. Softw. 1 (1975), no. 2, 108–128.

[2] Webb Miller and David Spooner, *Software for roundoff analysis*, ii, ACM Trans. Math. Softw. 4 (1978), no. 4, 369–387.

[3] Webb. Miller and Celia Wrathall, *Software for roundoff analysis of matrix algorithms*, Academic Press, New York, 1980

Performance Evaluation of Storing Inhomogeneous Descriptive Data of Digital Maps

Roberto Giachetta and Zsigmond Máriás

In contemporary Geographical Information Systems (GIS) the large variety of digital map sources requires the handling and storing of various descriptive data in a single storage facility. When even the number and type of attributes can vary map by map, the storage of such inhomogeneous data in a single database is difficult, as querying is an essential task and requires fast retrieval of data based on any present attribute. Examples of such databases can be found under the domain of the Institute of Geodesy, Cartography and Remote Sensing.

This paper presents a performance analysis study regarding the storage of descriptive data of digital map in database systems. Map description data can vary from simple numbers and text to complex data structures, like coordinates and timelines. Even with creating different categories for maps, it cannot be guaranteed that all items in the category have the same types of attributes stored. In traditional relational databases, tables must have a predefined structure according to the database schema, therefore simply providing columns for each descriptive data this kind of flexibility cannot be met, and data must be grouped and stored in given columns. Therefore fast querying features of relational databases (including indexing) cannot be used, which results in significant performance loss. The authors have studied several solutions addressing this problem.

The first solution is the usage of a semi-structured document-oriented database system called MongoDB, which defines database items as documents that can contain any kind and any number of attributes (even other documents) regardless of any predefined structure. These items are indexed according to their attribute types and values. These documents can be stored in collections, and there is no restriction on the structure of documents within a collection, therefore no restriction is needed on descriptive data.

The second solution is based on the creating an object-oriented environment inside a relational data model (based on our previous studies) and storing digital maps as objects in the database. This structure relies on creating inheritance taxonomy within the database and providing facilities to create classes and objects that can be altered in run-time. We have two approaches for this solution. In the first one, attributes' descriptions and classes are described in separate tables with a third table providing the link between a class and its attributes. Objects are stored in one table, and different types of attribute data are stored in different tables. This model has the advantage of being very flexible, when changing class structures with no need to redefine any database schema, however, queries must fetch object data using different tables, and inheritance structure is hard to reconstruct. Several improvements have been introduced for these problems (denormalization, inherited field storage, attribute table contraction). In the second approach class and attribute schemas are stored in the same way, but a separate table is created for each class to store the instances. Therefore creating, altering or even dropping tables is needed when changing schemas, but querying will be much faster.

In this study, the authors compare the original rational database solutions with these three approaches by implementing and testing with massive inhomogeneous and altering descriptive data. The article will conclude by presenting which method is best with digital map data at different search/insert/update command ratios. Since this problem is not only typical in the field of GIS, the solution can be applied generally to any domain using inhomogeneous data, like e-commerce systems and document warehouses.

Keywords: geographical information systems, large-scale spatial data storage, document-oriented databases, object-oriented databases, performace analysis.

Acknowledgements

The research is supported by the TÁMOP-4.2.1/B-09/1/KMR-2010-003 project.

References

[1] Csornai, G., László, et al., The Integrated Utilization of Satellite Images in Hungary: Operational Applications from Crop Monitoring to Ragweed Control, *Proceedings of the 26th Symposium of EARSeL, Millpress*, Rotterdam, The Netherlands, 2007.

[2] Chodorow, C., Introduction to MongoDB, *Free and Open Source Software Developers' European Meeting (FOSDEM)*, Brussels, Belgium, 2010.

[3] Giachetta, R., Elek., I., Developing an Advanced Document Based Map Server, *8th International Conference on Applied Informatics (ICAI)*, Eger, Hungary, 2010.

[4] Máriás, Zs., Design and Performance Analysis of Hierarchical Large-scale Inhomogeneous Databases, *8th International Conference on Applied Informatics (ICAI)*, Eger, Hungary, 2010.

[5] Nadkarni, P. M., et al., Organization of Heterogeneous Scientific Data Using the EAV/CR Representation *The Journal of the American Medical Informatics Association (JAMIA)*, vol. 6, issue 6, 1999.

Comparison of Clustering and Community Detection Algorithms

Erika Griechisch

The datamining became an important field in the last few years. Usually every single data (record) connects to a person, and people have relationships between each other. After determining these relationships, we can use graph representation to understand the connections in the background. In graph based datamining, we can determine the important points and the significant groups in networks.

In our study we examined and tested some of the clustering algorithms and community detection methods on a few small world graphs. The clustering algorithms which we investigated are the algorithm based on edge betweenness, the leading eigenvector method, the label propagation method, the greedy modularity maximalization method, and the spinglass method based on statistics. Moreover we present the results of two community detection algorithms: the clique percolation method and the algorithm based on N^{++} sets.

The modularity is an important measure of a particular division of network. We gave an extension of the modularity based on the definition of fuzzy partition matrix. Our aim was to maximize the extended modularity, because optimum determines a fuzzy partition matrix, where every row of the matrix represents the membership values of a node.

The optimum of the extended modularity on the three test graphs usually gave a strict partition, thus we got a binary fuzzy partition matrix. In some cases we got overlaps.

Median Filtering in Algebraic Reconstruction Methods

Norbert Hantos and Péter Balázs

The main task of image reconstruction is to create the 2D image of a cross-selection of an object from its projections. Algebraic methods can provide good solutions if only few projections are available. However, these methods – especially the Algebraic Reconstruction Technique – create noisy images if the iteration number or the computational time is limited.

In this presentation, we show how to decrease the noise in the resulted image using median filters during the reconstruction. Median filters are common tools in image restoration to decrease salt-and-pepper noise in a digital image without blurring the edges.

For testing we implemented the Algebraic Reconstruction Technique, the Simultaneous Iterative Reconstruction Technique (continuous reconstructions) and the Discrete Algebraic Reconstruction Technique (binary reconstructions). For median filtering we used Simon Perreault's implemented algorithm [1]. Our artificial test images contained homogeneous regions with different levels of topology. During the reconstruction we implicitly searched for an optimal parametrization of the filtering. We discuss the developed methods and present the experimental results, consequences and future plans as well.

References

[1] S. Perreault, P. Héber: Median filtering in constant time. *IEEE Transactions on Image Processing*, 16 (2007), 2389-2394.

Mathematical Models and Methodologies in Batch Process Scheduling

Máté Hegyháti

Batch process scheduling is one of the most important problems to be solved in process systems engineering, since it is of upmost importance for a chemical facility to utilize the available resources and equipment units in the most favourable way. In general, the aim of scheduling is the assignment of tasks to equipment units and time intervals in consideration of a certain objective.

Most of the work published in the literature applies mixed-integer linear programming (MILP) as the mathematical model to tackle these type of problems. The two main features, that usually characterize an MILP formulation is the selection of the binary variables and the representation of the time domain. A comprehensive overview of the available models has been published by Floudas and Lin [1] and Mendez *et. al* [2]. Even though literature provides many papers on the description of different formalizations, there are no studies considering and analysing the modelling step in the optimization process. As the validity of the models is usually not proven, unexpected behaviour can arise, which may result in suboptimal or even infeasible solutions of the original problem. A recent study by Hegyháti *et. al.* [3] has examined a literature example, where the supposedly optimal solution was infeasible in practice.

Sanmarti *et. al.*[4] has developed the S-graph framework for the algorithmic scheduling of batch chemical processes with non-intermediate storage policy. Unlike the MILP formulations, this new graph-theoretic methodology ensures the validity of the model, thus, it always provides the globally optimal solution. In later publications, the framework has been extended to handle a wider range of scheduling problems by Majozi and Friedler[5], and Adonyi *et. al.*[6, 7].

In the present work the published mathematical models and methodologies are detailed and analized, focusing on their validity, flexibility and efficiency.

References

[1] Floudas C.A. and X. Lin, "Continuous-Time versus Discrete-Time Approaches for Scheduling of Chemical Processes: A Review", Computers and Chemical Engineering, 28 (11), 2109-2129 (2004).

[2] Mendez, C.A., J. Cerdá , I. E. Grossmann, I. Harjunkoski, and M. Fahl, "State-Of-The-Art Review of Optimization Methods for Short-Term Scheduling of Batch Processes," Computers & Chemical Engineering 30, 913-946 (2006).

[3] Hegyháti, M., T. Majozi, T. Holczinger, and F. Friedler, Practical infeasibility of cross-transfer in batch plants with complex recipes: S-graph vs MILP methods, Chemical Engineering Science, 64, 605-610, (2009).

[4] Sanmarti, E., T. Holczinger, L. Puigjaner, and F. Friedler, Combinatorial Framework for Effective Scheduling of Multipurpose Batch Plants, AIChE Journal, 48(11), 2557-2570 (2002).

[5] Majozi, T. and F. Friedler, Maximization of Throughput in a Multipurpose Batch Plant Under Fixed Time Horizon: S-Graph Approach, Ind. Eng. Chem. Res., 45, 6713-6720 (2006).

[6] Adonyi, R., J. Romero, L. Puigjaner, and F. Friedler, Incorporating heat integration in batch process scheduling, Applied Thermal Engineering, 23, 1743-1762 (2003).

[7] Adonyi, R., G. Biros, T. Holczinger, and F. Friedler, Effective scheduling of a large-scale paint production system, Journal of Cleaner Production, 16 (2), 225-232 (2008).

Speeding up Browsing with Caching Regular Expressions

Renáta Hodován

It is clearly visible that the Web is becoming the most popular application platform. Static web pages fade into the past; more and more web sites use advanced techniques like server and client-side scripting, and AJAX to improve the user experience. For client-side scripting, JavaScript is by far the most commonly used dynamic language [6], it is supported by all major browsers. Not surprisingly, the optimization of JavaScript execution is a hot research topic [5,7]. However, as our experiments show, attention should be paid to regular expressions as well. In this work, we present a study on how regular expressions are used on popular web sites, introduce the idea of caching regular expressions, and present results of an experimental implementation.

First, the most popular web sites [1,2,4,8] have been collected in four categories (community, news, torrent, and adult) and then we visited those pages. With a modified version of WebKit [3] – a popular browser engine used in several desktop and mobile browsers – all regular expression patterns has been logged. Thus, we got a list of patterns representing the real-life load of a regular expression engine. Table 1 shows the number of regular expressions parsed and compiled during the browsing sessions, and the number of unique patterns as well. It is interesting to see how low the ratio of unique regular expressions is: less than 4%, irrespectively of the type of the visited pages.

Table 1: Regular expressions on web pages

	Community	News	Torrents	Adult
Total	50243	63165	22605	48666
Unique	1529(3.04%)	1972(3.12%)	889 (3.93%)	701 (1.43%)

The above results motivated us to experiment with caching. A cache has been implemented where the internal representations of the most recently compiled regular expressions are stored. Round-robin caching policy has been used with a cache size of 256. Table 2 shows the hit-miss ratio of the cache and Table 3 presents the effect of caching on the performance of regular expression parsing and compilation. (We experimented with the more complex Least Recently Used policy as well, but it was not significantly better.) As the latter Table shows, caching can result in a reduction in regular expression parsing and compilation time as high as 69%, which equals to a 3-fold speedup.

Table 2: Hit-miss ratio of the cache

	Community	News	Torrents	Adult
Hit	47240	28176	20575	46690
Miss	2963	3309	2030	1976

Table 3: The effect of caching on regex compilation time

	Community	News	Torrents	Adult
Original (ms)	354.862	190.309	170.395	197.131
Cached (ms)	124.826	73.417	52.088	86.809
Gain	65.19%	61.41%	69.43%	55.96%

References

[1] News Top 20. The pick of the best news sites on the net, May 2010.

[2] Inc Alexa Internet. Top sites by cathegory: Adult, May 2010. http://www.alexa.com/topsites/category/Top/Adult.

[3] Apple, Inc. TheWebKit open source project. http://webkit.org/ (Accessed 1 July 2009).

[4] eBizMBA. Top 15 most popular torrent websites, May 2010. http://www.ebizmba.com/articles/torrent-websites.

[5] Andreas Gal, Brendan Eich, Mike Shaver, David Anderson, David Mandelin, Mohammad R. Haghighat, Blake Kaplan, Graydon Hoare, Boris Zbarsky, Jason Orendorff, Jesse Ruderman, Edwin Smith, Rick Reitmaier, Michael Bebenita, Mason Chang, and Michael Franz. Trace-based just-in-time type specialization for dynamic languages: *In Proceedings of the 2009 ACMSIGPLAN conference on Programming language design and implementation (PLDI'09), pages 465-478, Dublin, Ireland, June 2009.*

[6] Zoltán Herczeg, Gábor Lóki, Tamás Szirbucz, and Ákos Kiss: Guidelines for JavaScript programs: Are they still necessary? *In Proceedings of the 11th Symposium on Programming Languages and Software Tools (SPLST'09) and 7th Nordic Workshop on Model Driven Software Engineering (NW-MODE'09), pages 59-71, Tampere, Finland, August 26-28, 2009. Tampere University of Technology.*

[7] Maciej Stachowiak: Introducing SquirrelFish Extreme, September 2008. http://webkit.org/blog/214/introducing-squirrelfish-extreme/ (Accessed 1 July 2009).

[8] Wikipedia. List of social networking websites, May 2010. http://en.wikipedia.org/wiki/List_of_social_networking_website.

New Methods in Payload Based Network Traffic Classification

Béla Hullár, Sándor Laki, András György, and Gábor Vattay

The services provided over the Internet have gone through an enormous evolution in the last decade. Numerous new services and applications have emerged (eg. VoIP, IPTV, file sharing, etc). The appearance of these new services and the growing number of network attacks have brought along the development of new traffic analyzer methods. Network operators would like to identify and control the traffic that travels through their networks. This knowledge may lead to more efficient resource allocation strategies and improved service quality.

In the past different services used their well known TCP or UDP port numbers defined by the Internet Assigned Numbers Authority (IANA). Nowadays most of the applications use dynamic port numbers or well-known trusted ports (such as HTTP or SMTP ports). The reason of this behavior is that applications have to by-pass firewalls and routers, while others try to hide their presence. State-of-the-art, widely used traffic analyzer applications are mostly based on the, so called, deep packet inspection (DPI) approach that aims at identifying typical protocol-patterns in the messages of different network applications. Since these methods cannot handle encrypted traffic, recently several machine learning based traffic classification methods have been developed that do not consider the content of the packets [1]. Furthermore, although the majority of network traffic is still unencrypted and the DPI solutions perform well in practice, their disadvantages are well-known: these methods require significant computing resources and the integration of new applications requires expert knowledge about the application's protocol. Statistical payload inspection can give a proper solution to these issues [2, 3].

This paper examines how different data compression models can be used for packet payload based protocol identification. Our results show that compression-based modeling can provide an effective solution for traffic classification, similarly to their performance in other domains of clustering [4]. We found that the majority of the protocols are identifiable from the first several bytes of the application data. To demonstrate our results, different real network traces generated by different network applications were used.

Acknowledgements

The authors thank the partial support of the National Office for Research and Technology (NAP 2005/ KCKHA005) project.

References

[1] Thuy T.T. Nguyen, Grenville Armitage: A Survey of Techniques for Internet Traffic Classification using Machine Learning, *IEE Communications Surveys and Tutorials, vol. 10 no. 4 pp. 56-76, 2008.*

[2] A. Finamore, M. Mellia, M. Meo, and D. Rossi: KISS: Stochastic Packet Inspection,*In Traffic Measurement and Analysis (TMA), Springer- Verlag LNCS 5537, May 2009.*

[3] Justin Ma, Kirill Levchenko, Christian Kreibich, Stefan Savage, Geoffrey M. Voelke: Unexpected means of protocol inference, *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement, October 25-27, 2006, Rio de Janeriro, Brazil*

[4] R. Cilibrasi, P.M.B. Vitanyi: Clustering by compression,*IEEE Trans. Information Theory, 51:4(2005), 1523-1545.*

Guided Exploration in Policy Gradient Algorithms with Gaussian Process Function Approximation

Jakab Hunor

Gradient based direct policy optimization algorithms have strong convergence properties and can be used for model-free learning of control policies in complex high-dimensional Markov Decision Processes. The main drawback of the majority of Policy Gradient methods however is the high variance of the gradient estimate which is a result of approximating the Q-values in the gradient expression by Monte Carlo sampling of rewards accumulated through individual trajectories $Q^{\pi}(x,a) \sim \left(\sum_{j=o}^{H} \gamma^j r_j\right)$. In [1] we have presented a method for reducing the gradient variance by using a Gaussian Process based function approximator on state-action space to replace the high-variance(although unbiased) Monte Carlo estimation $Q(\cdot,\cdot) \sim GP(m_q, k_q)$. In this paper we present several ways to extend this method by exploiting the fully probabilistic nature of the Gaussian Process estimates to influence the directions of exploration. As we are mainly interested in robotics related learning problems, it is essential for us to develop methods that can be performed online. Also because of the nature of these learning problems, the number of trials that can be performed during the learning period, and the regions of the search-space that can be reached are limited both by physical and time constraints. Therefore random exploration in our methods performs poorly and needs to be replaced with guided exploration. In basic versions of Policy Gradient algorithms exploratory behavior is introduced by constructing the action-selection policy π from a deterministic controller output and an added exploratory noise: $\pi(a\|s) = f(s,\theta) + e$ where $e \sim N(0,\sigma^2)$. We replace the exploratory noise e by taking into account the GP predictive variance $k_q(s,a)$ at the state-action pair (s,a) which gives us useful information about how well the current region of the search-space has been explored. We design our search criteria based on the concept of exploitation and exploration. At each time-step for a given state s our controller returns a specific action $f(s,\theta) = a$. We sample a number of points from the neighborhood of a and get the predicted Q-values and variances for these points paired with state s. At this point we can choose to select an action from the sampled points which has the highest predicted Q-value (exploitation) or one that has the highest predictive variance (exploration). We define a parameterized measure which balances between these two. The exploratory noise's mean and variance will be set in such a way as to shift the policy π in the direction of the action selected in the above described way. We also investigate a simplistic approach where we only change the exploratory noise's variance from fixed σ^2 to the predictive variance of the selected state-action pair (s,a). We test the proposed methods on the inverted pendulum simulated control task in MATLAB where we compare its performance to our previous algorithm. We also perform tests on a pole-balancing control problem in a realistic simulation environment (ODE) to verify the behavior of the algorithm in uncertain environments.

Acknowledgements

This work was supported by programs co-financed by The SECTORAL OPERATIONAL PROGRAMME HUMAN RESOURCES DEVELOPMENT, Contract POSDRU 6/1.5/S/3 - "Doctoral studies: through science towards society".

References

[1] L. C. Hunor Jakab: Using Gaussian processes for variance reduction in policy gradient. *In ICAI2010: Proceedings of the 8th International Conference on Applied Informatics*, 2010.

An Ideal Anonymous E-exam System

Andrea Huszti and Péter Pleva

As electronic assessment forms an essential part of e-learning environments and, ideally, trustworthiness is of key importance in such systems, the design principles of e-exam schemes must be considered carefully. We propose a protocol that, above all the basic security requirements that traditional paper-based exams meet (such as secrecy, correctness and authenticity), provides anonymity for both examinees and examiners. From examinees' point of view, being anonymous not only ensures objectivity in determining their grades but also prevents partiality. On the other hand, examiners are also protected against bribing and threatening attacks with the help of anonymity.

Participants of our scheme include examinees, examiners, an Examination Board and an Administration Authority but the involvement of a Certification Authority and Timestamp Service Provider is also necessary. Examinees and examiners register for taking and correcting the exam, respectively, at Administration Authority. Examination Board receives e-exams and forwards them to examiners chosen randomly. Certification Authority and Timestamp Service Provider is responsible for managing digital certificates, issuing timestamps and also for controlling anonymous servers that replace senders' IP addresses to their own ones and perform some cryptographic operations. We assume honest behaviour on the part of the Certification Authority and Timestamp Service Provider, solely.

Acknowledgements

This work is supported by the national Economic Operative Programme (GOP-2007-1.1.2).

References

[1] Jordi Castella-Roca, Jordi Herrera-Joancomarti, Aleix Dorca-Josa: A Secure E-Exam Management System, ares, pp.864-871, First International Conference on Availability, Reliability and Security (ARES'06), 2006

[2] Andrea Huszti, Attila Pethö: A Secure Electronic Exam System, to appear

[3] ExamSoft Worldwide, `http://www.examsoft.com`

Algorithmic Synthesis of Structural Alternatives for Business Processes

Károly Kalauz, Botond Bertók, and Ferenc Friedler

Design and management of business processes are key factors for companies to be competitive. Available computer aid for business process analysis and optimization can help evaluating decision alternatives if available. In contrast, no software in the market supports exploring or generating structural alternatives for business processes. A standard Business Process Modeling Notation (BPMN) provides businesses the capability of understanding their internal business procedures in a graphical notation and gives organizations the ability to communicate these procedures in a standard manner. BPMN defines a Business Process Diagram (BPD), which is based on a flowcharting technique tailored for creating graphical models of business process operations. In the last decades, several robust and reliable process network optimization algorithms have been developed and implemented on the basis of the P-graph framework (Friedler *et* al.). P-graph is the only approach for process-network optimization which provides algorithmic mathematically proven solution for its each step, i.e., superstructure generation, construction of the mathematical model, optimization, and the solution interpretation. Appropriate adaptation of these algorithms would be beneficial for business process design and optimization. The application of the P-graph framework for business process optimization requires the algorithmic reformulation of the available business process models in the form of P-graphs and adaptation of the algorithms to the special structural properties of the business processes. The aim of our research is to develop a methodology that provides an adequate basis to describe and model business processes, as well as to algorithmically synthesize optimal and alternative business processes. Widely used business process modeling notations and theoretical results of algorithmic process-network synthesis are to be integrated. A novel approach for BP optimization has been proposed based on the P-graph methodology. Model transformation of BPDs to P-graphs as well as algorithms for generating structural alternatives for business processes have been elaborated and implemented. The proposed methodology has been applied to practical problems of recent R&D projects.

Order-Independent Sequential Thinning on Two-Dimensional Binary Images

Péter Kardos

The aim of skeletonization is to extract a region-based shape feature, the medial curve of binary objects. Blum illustratively interpreted this process as the fire front propagation [1]. Thinning is a frequently used strategy for this purpose. Thinning algorithms iteratively peel off points from the boundary of the objects which satisfy some geometrical and topological conditions [2]. Although the fire front propagation is parallel by nature, several sequential thinning algorithms were proposed, as well [3]. The motivation behind the use of the latter methods is that topology preservation can be easily guaranteed when only one non-skeleton point is removed at a time. However, sequential thinning algorithms have also a major weakness: they may produce various medial curves for different visiting orders of border points.

Earlier we published an order-independent sequential algorithm which is based on a classification of border points [4]. In the beginning of an iteration the algorithm labels these special border points, and later it takes advantage of this additional information to decide whether a point is deletable or not.

Because of the use of extra labels one may say that such an algorithm cannot be considered as "sequential" in a strict sense. This motivated us to define another thinning scheme, which considers tricolor input images in which background points, boundary object points, and non-boundary object points are indicated by different colors, and no additional labeling is allowed during the process. In this work, we propose a template-based order-independent sequential thinning algorithm that is based on this scheme.

References

[1] Blum, H.: A transformation for extracting new descriptors of shape. Models for the Perception of Speech and Visual Form, MIT Press, 1967, 362–380

[2] Kong, T.Y., Rosenfeld, A.: Digital topology: Introduction and survey. Computer Vision, Graphics, and Image Processing **48**, 1989, 357–393

[3] Lam, L., Lee, S.-W., Suen, C.Y.: Thinning methodologies — A comprehensive survey, IEEE Trans. Pattern Analysis and Machine Intelligence **14**, 1992, 869–885

[4] Kardos, P., Németh, G., Palágyi, K.: An Order-Independent Sequential Thinning Algorithm. 13th International Workshop, IWCIA 2009, LNCS 5852, pp. 162-175

Employing Pythagorean Hodograph Curves for Artistic Patterns

Gergely Klár and Gábor Valasek

In the paper we will present our design element creator tool for the digital artist. The purpose of our tool is to support the creation of vines, swirls, swooshes and floral components. To create gentle and pleasing curves we employ Pythagorean hodograph (PH) quintic curves to join a hierarchy of *control circles* defined by the user. The control circles are joined by spiral segments with at least G^2 continuity, ensuring smooth and seamless transitions. The control circles give the user a fast and intuitive way to define the desired curve. The resulting curves can be exported as cubic Bézier curves for further use in vector graphics applications.

Introduction and Background

Floral elements, vines, tangled spirals and similar features are among the most popular design components. These components could be found in traditional ornamental and contemporary abstract designs as well. There have been several efforts to automate aspects of the artistic process of creating such designs and ornaments. Since the fundamental work of Prusinkiewicz and Lindenmayer [1] L-systems has been used widely to generate flowers and flower like patterns. Wong et al. [2] presented a system to automatically generate space-filling floral ornaments. Their system uses proxy objects during generation what could be replaced by arbitrary elements created by any means. Xu and Mould's *Magnetic Curves* [3] are more closely related to our work. They focus on the creations of the curves themselves, but their method uses a discrete time-step approach, and they commit the problem of creating smooth curves to approximation routines.

We based our work on the same assumption as Xu and Mould, that is, a pleasing curve has a smooth curvature. Our goal is to make a tool for the digital artists to create such curve or sequence of curves with ease. Out method uses cubic and quintic splines to generate resolution independent curves, which can be used themselves, serve as a skeleton of a design, or act as a path for strokes or objects. To support the widest range of third party tools possible, the generated curves can be exported as cubic Bézier splines.

PH curves introduced by Farouki and Sakkalis [4] have very favourable properties, most importantly it is possible to define spiral segments using PH quintic curves whose curvature changes monotonically with arc-length. These curves have been used in highway, railway and robot trajectory design. Now we would like to show the efficiency of these curves as design elements as well.

Our tool In our tool the user defines a hierarchy of control circles to create her design. If it is possible the system automatically connects a circle to its ancestor with an appropriate curve.

Our system supports circle-to-circle S-shaped curves, circle-inside-circle transitional spiral segments, and circular arcs approximated by Bézier curves joining the two preceding. During design additional properties are specified for each control circle.

Both for the circle-to-circle S-curves and circle-inside-circle spiral segments Pythagorean hodograph quintic curves are used. These curves have been defined to have G^2 contact to their control circle. Therefore, each incoming curve can have G^2 contact with any outgoing curve of the same circle. Two non-touching, non-overlapping control circles are connected by an S-shape, if within a certain threshold. The derivation for the control points of the S-curves is following the work of Walton and Meek. The positions and radii of the circles define the shape of the S, the only additional user input required is whether the shape should be mirrored or not.

A fully contained circle is joined to its ancestor with a spiral segment, if such transition is possible. The conditions and the derivation of the control points are given in Habib and Sakai's [5] work. The radii of the circles define the range of allowed distances between the centres. For now, our software chooses the smallest possible distance. For a given pair of radii and distance, the spiral segment is uniquely defined, thus the radii and centre-to-centre distance dictates the

positioning of the smaller circle. If the creation of the transition curve is possible, only the radius of the smaller circle is used, and its centre is repositioned as defined by the algorithm. Similarly to the S-curves, a mirroring property can be defined. An additional parameter of a spiral segment is its starting point, defined in degrees on the arc. For new segments, this is automatically calculated, so the new curve continues the ancestor's incoming curve, if it exists.

Because the computations of the control points in both cases are quite involved, especially for the second case, several numerical algorithms were required in the implementation. In the derivation of the curve parameter theta for the first case, Halley's method is used, since the simpler Newton's method proved to fail on several occasions. However, Newton's method is sufficient for the second case, during the calculation of the valid range of centre distances. Unfortunately, none of these methods were useable for determining the theta value in this case. The derivative of the concerned function is too complex to be useable. Nevertheless, because the domain of the possible solutions is known, a simple bisection method has proven adequate.

The conversion from Pythagorean hodograph quintic to cubic Bézier are done by elementary degree reduction, with endpoint correction. This is sufficient, because PH quintic curves have very similar geometric properties to cubic Béziers.

References

[1] Prusinkiewicz, Przemyslaw and Lindenmayer, Aristid: The algorithmic beauty of plants. Springer-Verlag New York, Inc., 1996.

[2] Wong, Michael T. and Zongker, Douglas E. and Salesin, David H.: Computer-generated floral ornament. SIGGRAPH'98: Proceedings of the 25th annual conference on Computer graphics and interactive techniques, 423-434, 1998.

[3] Ling Xu and David Mould: Magnetic Curves: Curvature-Controlled Aesthetic Curves Using Magnetic Fields. pages 1-8, Victoria, British Columbia, Canada, 2009. Eurographics Association.

[4] R. T. Farouki and T. Sakkalis. Pythagorean hodographs. IBM J. Res. Dev., 34(5):736-752, 1990.

[5] Zulfiqar Habib and Manabu Sakai: On ph quintic spirals joining two circles with one circle inside the other. Computer-Aided Design, 39(2):125-132, 2007.

Skinning of Circles Using Biarcs

Roland Kunkli and Miklós Hoffmann

Interpolation of points is a momentous problem in Computer Aided Geometric Design (CAGD). We have several (now standard) methods to solve this kind of problem [1]. If we consider circles instead of points, the interpolation will be transferred to "skinning". By skinning we mean the geometric construction of two curves (often G_1 or C_1 continuous) touching each of the circles and provide a result, which is visually "satisfactory". It is very hard to define this problem with mathematical formulae, and we have not got precise definition yet. Furthermore, we have not got a well-known, ubiquitously method to get a solution. With the help of skinning we can easily design tubular structures, and it can be useful by covering problems too. There is recent numerical approach of the problem [2], but it has a lot of weaknesses by certain positions of the circles. We have submitted a manuscript [3] to the journal *Computer Aided Geometric Design*, in which we present a method. With this presented algorithm we can easily compute skinning curves and it works in the above-mentioned problem cases too.

In this paper we present a new method, where the skinning curves are composed of *biarcs*. A biarc is a G_1 continuous curve, which is composed of two arcs [4]. The real power of this method is, that it uses only biarcs for the construction and it also works in the above-mentioned problem cases too. So it can be useful by applications, where this is essential, e.g. CNC cutting. It can easily implemented and it can be used by an interactive, dynamic software, because it is fast. Furthermore, the user must determine only the circles.

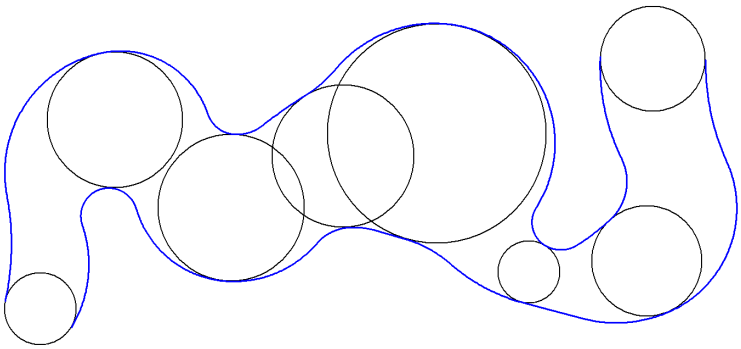


Figure 1: Skinning of circles using biarcs

Keywords: skinning, circles, interpolation, biarcs

References

[1] G. Farin, Curves and Surface for Computer-Aided Geometric Design, 4th edition, *Academic Press*, New York, 1997.

[2] G. Slabaugh, G. Unal, T. Fang, J. Rossignac, B. Whited, Variational Skinning of an Ordered Set of Discrete 2D Balls, *Lecture Notes on Computer Science*, 450–461, 2008.

[3] R. Kunkli, M. Hoffmann, Skinning of circles and spheres, *Computer Aided Geometric Design*, submitted.

[4] K. M. Bolton, Biarc curves, *Computer–Aided Design*, 7, 89–92, 1975.

[5] Z. Šír, R. Feichtinger, B. Jüttler, Approximating curves and their offsets using biarcs and Pythagorean hodograph quintics *Computer–Aided Design*, 38, 608–618, 2006.

Feature Selection and Classification Algorithms for Retinal Microaneurysm Detection

István Lázár

Diabetic retinopathy (DR) is a disorder of the retina, caused by diabetes mellitus (usually simply referred to as diabetes). Color fundus images are the main resources used by ophtamologists for screening purposes. Red lesions, such as microaneurysms (MAs) and haemorrhages are one of the first signs of DR, thus their reliable detection is essential in a computer aided screening system. Literature on automatic MA detection is extensive, the first proposed method was published in 1984 [1]. In [2] Niemeijer et al. compares the results of state of the art MA detectors in an online challenge that focuses on the automated detection of retinal disease. Currently available methods divide the problem of MA detection into two consequent stages. In the first stage, MA candidates - regions of the image, that probably correspond to MAs - are extracted using some specific image segmentation method. In the second stage (detection), first, each candidate is mapped into an n-dimensional feature space, resulting in a set of feature vectors. Elements of the feature vectors attempt to describe specific attributes of the candidates, that may help to separate true MAs from false candidates. Next, supervised machine learning (classification) methods are used to separate the feature vector set, denoting the candidates corresponding to vectors as true and false. Manually marked images are used to obtain the training vectors for the classification methods. At the present time, several publicly available databases exist for training and testing MA detection algorithms [3].
In this paper, we examine the problem of optimal feature subset selection, and survey the usage of classification algorithms, as described in many proposed MA detection methods. So far, more than 50 features have been proposed for binary classifiers in MA detectors. Some features are rather simple and extensively used, such as the circularity or mean intensity value of the candidate region, and some others are strongly connected to a specific candidate extraction method. To find the optimal feature subset for a classifier, testing of all possible feature subsets is required. If large number of features are available, this approach is impractical. As a matter of fact, it is technically impossible in this specific case. Two typical categories of feature selection are subset selection and feature ranking. In the case of subset selection, a search algorithm is used to scan through the space of possible subsets, and the current subset is iteratively modified based on some scoring metric. Then, the new subset is evaluated as a group, and on the basis of the new subset's suitability, the algorithm decides on terminating or continuing the search. Such search approaches are simulated annealing and genetic algorithms. Feature ranking is based on ranking each feature individually by a metric, and eliminating the ones that do not achieve a sufficient score. Popular scoring metrics are e.g. correlation, entropy, mutual information, or Kullback-Leibler divergence (information gain). The optimal feature subset depends not only on the used classifiers, but also on the candidate extraction method, i.e. different feature set is optimal for every candidate extractor, classifier pair. Classification methods used for the final decision in MA detectors include e.g., nearest neighbour algorithms, artificial neural networks, support vector machines, and kernel density estimators.
As a result, a new feature set is proposed, with better performance in the previously mentioned classifier methods for two publicly available datasets from [3]. The quantitative comparison of the performance of the binary classifiers is performed in terms of ROC and FROC curves [4].

Acknowledgements

This work was supported in part by the János Bolyai grant of the Hungarian Academy of Sciences, and by the TECH08-2 project DRSCREEN - Developing a computer based image processing system for diabetic retinopathy screening of the National Office for Research and Tech-

nology of Hungary (contract no.: OM-00194/2008, OM-00195/2008, OM-00196/2008). We also acknowledge the Moorefields Eye Hospital, London for their clinical support.

References

[1] C. E. Baudoin, B. J. Lay, and J. C. Klein: Automatic detection of microaneurysms in diabetic fluorescein angiographies. *Revue D' Épidémiologie et de Sante Publique,* vol. 32, pp. 254-261, 1984.

[2] M. Niemeijer, B. van Ginneken, M. J. Cree, A. Mizutani, G. Quellec, C. I. Sanchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, X. Wu, G. Cazuguel, J. You, A. Mayo, Q. Li, Y. Hatanaka, B. Cochener, C. Roux, F. Karray, M. Garcia, H. Fujita, and M.D. Abramoff: Retinopathy Online Challenge: Automatic Detection of Microaneurysms in Digital Color Fundus Photographs. *IEEE Transactions on Medical Imaging*, vol. 29(1), pp. 185-195, 2010.

[3] M. Niemeijer, M.D. Abramoff and B. van Ginneken: Publicly available retinal image data and the use of competitions to standardize algorithm performance comparison. *IFMBE Proceedings*, vol. 25/11, pp. 175-178, 2009

[4] P. Bunch, J. Hamilton, G. Sanderson, and A. Simmons: A free response approach to the measurement and characterization of radiographic-observer performance. *J Appl Photogr Eng*, vol. 4, pp. 166-172, 1978.

Distributing Agent-Based Simulation

Richárd O. Legéndi and Attila Szabó

The exponential growth of computational capacity of personal computers played an important role in the spread of computational modeling. Simulation techniques such as Agent-Based Modeling (ABM) are becoming a common practice, where enormous runs claim vast computational resources because of the need of precise and robust large scale experimental results. Multi-core systems (like clusters, grids, clouds, or even desktop PCs) offer multiplied resources at a relatively low cost, but effective utilization of these systems is still a challenge for software technology.

In the context of ABM distribution, the dynamics of the communication network between agents play an important role. When the parts of a distributed model are highly interconnected, the overhead of network communication can easily reduce the performance of the multi-core system. Gulyás et al. defined a classification of six types (based on common communication schemes), and introduced a solution to distribute models having static communication networks [1]. Scheutz and Schermerhorn defined adaptive algorithms for parallel execution of "spatial AMBs", where agents move in a spatial environment [2]. Gilbert et al. used a peer-to-peer infrastructure to simulate the emergence of artificial societies [3]. Yamamoto et al. created a simulation environment that enables to host millions of agents on a system of workstations connected with a high performance network [4].

In this paper we study the performance of some existing tools of distributed software execution by running agent-based simulations on a cluster of computers. We measure the effectiveness and practical scalability of these frameworks by implementing selected models available at the the OpenABM Consortium's on-line model repository [5] using ProActive [6], JavaSpaces [7], and other Java RMI-based tools [8]. We also seek for effective implementation techniques and practices to support large-scale simulations, those exceeds the limitations of a single computer (e.g. by using several gigabytes of memory). Our goal is to reach a competitive simulation speed compared to sequential execution by keeping the cluster's communication overhead at a relatively low level.

Acknowledgements

This work was supported by a grant from the Hungarian Government (TÁMOP-4.2.1/B-09/1/KMR-2010-0003).

References

[1] Gulyás, L. et. al.: Templates for Distributed Agent-Based Simulations on a Quasi-Opportunistic Grid, *EMSS* (2008).

[2] Scheutz, M., and Schermerhorn, P.: Adaptive Algorithms for the Dynamic Distribution and Parallel Execution of Agent-Based Models, *Journal of Parallel and Distributed Computing*, 66, 1037-1051., (2005).

[3] Gilbert, N. et al.: Emerging Artificial Societies Through Learning, *Journal of Artificial Societies and Social Simulation 9* (2006).

[4] Yamamoto, G., Tai, H., and Mizuta, H.: A platform for massive agent-based simulation and its evaluation, *In Proceedings of the 6th international Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '07)* (2007).

[5] Model Archive, *OpenABM Consortium's site. Web* (7. May, 2010.)

[6] Baude, F. et al.: GRID COMPUTING: Software Environments and Tools, *Jose C. Cunha and Omer F. Rana (Eds), Springer Verlag* (2006).

[7] Freeman, E., Arnold, K., and Hupfer, S.: Javaspaces Principles, Patterns, and Practice, *1st. edition Addison-Wesley Longman Ltd.* (1999).

[8] Wollrath, A., Riggs, R., and Waldo, J.: A Distributed Object Model for the Java System (2010).

Analysis of ODF Document Schema Transformations

Zsolt Lengyel and Sándor Sike

Open Document Format (ODF) is an office document format used by several desktop office suites. Our project focuses on providing support of editing ODF-based documents on mobile devices. The limited capacity of mobile devices compared to desktop computers requires that documents should be simplified to be able to handle on mobile devices.

The basis of document simplification is the transformation of the original ODF document schema. The new schema defines the needed restrictions on the documents to be used on mobile devices. The schema transformation has to be properly specified and the resulting schema has to be checked against the specification. The grammar defined by the original ODF document schema is too large and complex to be handled manually, and so are the transformed schemata, therefore automatic tool and method is necessary for specification and verification. Schema transformations can be composed by particular sequences of schema transformation primitives. The specification of these primitives is relatively simple and their effect can also be described.

We introduce a formal model to express the method of schema transformations in a comprehensive way and for verification the properties of the resulting schema.

Acknowledgements

Supported by NKTH under TECH_08-A2/2-2008-0089. Special thanks to Zoltán Horváth and the SZOMIN team.

References

[1] Barna, I., Bauer, P., Bernád, K., Hernáth, Zs., Horváth, Z., Kőszegi, B., Kovács, G., Kozsik, T., Lengyel, Zs., Roth, R., Sike, S., Takács, G.: ODF Mobile Edition – Towards the development of a mobile office software, to appear in Proceedings of ICAI 2010 – 8th International Conference on Applied Informatics this year

[2] Barna, I., Bauer, P., Bernád, K., Hernáth, Zs., Horváth, Z., Kőszegi, B., Kovács, G., Kozsik, T., Lengyel, Zs., Roth, R., Sike, S., Takács, G.: A client-server model for editing ODF documents on mobile devices, submitted to CSE 2010

Calculating Non-Equidistant Discretizations Generated by Blaschke Products

Levente Lócsi

In many cases non-equidistand discretizations (or non-uniform divisions) have been proven very useful. Many examples can be found from the fields of computer graphics to FFT analysis by engineering sciences.

In [1] we introduced a very elegant way of handling non-uniformity in the case of signals (e.g. ECG signals) with regions of high variability and therefore more detail, dense discretization needed, and with constant-like regions where less detail, sparse discretization is enough. The Blaschke functions, Blaschke products and their associated argument functions are used to describe a suitable non-equidistant discretization. The inverse image of an equidistant discretization according to an argument function is considered.

One can give an explicit form of the inverse of an argument function associated to a Blaschke function: the inverse can be simply calculated. But in the case of Blaschke products, the inverse of the argument function has no explicit form, numerical methods are needed to solve the arising non-linear equations. We have as many equations as the number of points in the discretization to generate.

In the work to be presented here we analyse the efficiency of methods like the well-known bisection method and Newton's method applied to this problem. By taking advantage of non-uniformness, we may solve the equations at hand in a clever order also to be explained. The advantages and disadvantages of these methods and their combinations are to be analysed.

References

[1] L. Lócsi. Discrete approximation of ECG signals, *International Conference on Applied Informatics 2010*, submitted.

Active Control of Cable Bridges Using Fuzzy Logic

Shaghaghi K. Manouchehr

The response of a high-rise building and Suspension Bridge to dynamic forces such as earthquake and wind loads has been of primary interest to civil engineers. In the last two to three decades, control devices, passive as well as active, have been developed to suppress structural vibration due to these environmental disturbances. Among the concepts behind the development of these devices. The one based on the use of a mass as an added energy-absorbing system has been under rather intensive study, and the results have been fruitful (MCNamara 1977). Luft (1979), Ayorinde and Warburton (1980) presented approximated formulas for the optimal parametric design of a tuned-mass-damper (TMD) system.

In an attempt to increase the effectiveness of a TMD system, Chang and Song (1980) introduced an active force to act between the structure and the TMD system. The active systems in recent years been extensively taken into attention. Between the active controlling systems method can be called the ATMD method. To date, various strategies have been proposed for determining the active control force that some methods may be used in some structures.

A process for designing an effective active-tuned-mass damper (ATMD) system to control a tall building and Suspension Bridge subjected to stationary random wind forces was proposed by Abdel-Rohman (1984) using the pole- assignment method. The results suggested that the design of an optimal ATMD required at least a parametric study to select the ATMD parameters In this study, to determine the optimal ATMD active force control system we used the LQR algorithm and fuzzy controller. The structure is a Suspension bridge that its vertical vibration under earthquake loads in two modes, without control and controlled is studied.

Keywords: Active Tuned mass damper (ATMD), Tuned mass damper (TMD), Fuzzy Logic controller, LQR algorithm, Suspension Bridge

The Multiobjective Optimization Package of IOSO; Applications and Future Trends

Amir Mosavi

Indirect Optimization on the basis of Self-Organization (IOSO), first introduced by Professor Egorov I.N., designed for solving complex problems faster, has been successfully applied in searching for optimal decisions. It is based on the Response Surface (RS) methodology approach and on universal mathematical algorithms that can easily be applied to deal with MOO problems.

Based on IOSO it is assumed that if a problem could be represented by a mathematical model, IOSO optimization technology is able to approximate it into certain degrees. During operation, the information about the system behavior is stored for the points in the neighborhood of the extremum, therefore the RS model of design space will be more accurate providing wider range of capabilities, and would be practically insensitive with respect to the types of objective function and constraints.

In this paper the applications of IOSO in the field of industrial optimization are reviewed. This paper identifies recent approaches to utilizing the IOSO and the challenges that it presents to the Multiobjective Optimization (MOO) and Multiple Criteria Decision Making (MCDM) community. A survey of existing work, organized by application area, forms the main body of the review, following an introduction to the key concepts in MOO and MCDM. An original contribution of the review is the identification of strengths and weaknesses of technique, i.e. these are used to explain the reasons behind the use of IOSO in each application area and also to point the way to potential future uses of the technique. An overview of different approaches for Multiobjective design optimization utilizing IOSO is presented. The study identifies time as the major challenge and key to future success for IOSO strategy.

References

[1] Nomoto, H., Konga, A., Ito, S., Fukuyama, Y., Otomo, F., Shibuya,S., Sato, M., Kobayashi, Y. and Matsuzaki, H.: The Advanced Cooling Technology for the 1500 C Class Gas Turbine: Steam-Cooled Vanes and Air-Cooled Blades. *ASME Journal of Engineering for GasTurbines and Power*, Vol. 119, pp. 624-632. 1997.

[2] Dennis, B. H., Egorov, I. N., Sobieczky, H., Dulikravich, G. S. and Yoshimura, S.: Parallel Thermoelasticity Optimization of 3-D Serpentine Cooling Passages in Turbine Blades, ASME paper GT2003-38180, ASME Turbo Expo 2003, Atlanta, GA, June 16-19, 2003.

[3] Egorov, I. N.: Indirect Optimization Method on the Basis of Self-Organization", Curtin University of Technology, Perth, Australia., Optimization Techniques and Applications (ICOTA'98), Vol.2, pp. 683-691. 1998.

[4] Egorov, I. N., Kretinin, G. V., Leshchenko, I. A., Kostiuk, S.S: The Methodology of Stochastic Optimization of Parameters and Control Laws for the Aircraft Gas- Turbine Engines Flow Passage Components ", ASME paper 99-GT-227, 1999.

Algorithms for Topological Kernel Extraction and Thinning

Gábor Németh and Péter Kardos

A parallel reduction operation on binary images deletes a set of object points. There exist sufficient conditions for parallel reduction operations to preserve topology [3]. They provide a general method of verifying the correctness of topological algorithms.

There are two frequently applied topological algorithms. Reductive shrinking [1] is to extract topological kernels (i.e., minimal structures that are topologically equivalent to the original objects), and thinning that is a skeletonization technique [2]. All iterative parallel shrinking and thinning algorithms are composed of parallel reduction operations. The endpoints, some black points that provide relevant geometrical information with respect to the shape of the object, are preserved during the thinning process, in contrast to the reductive shrinking, where no endpoint criterion is taken into consideration.

In this work, we present new sufficient conditions for topology preservation. Then we propose 21 new algorithms, 7 for reductive shrinking and 14 for thinning. These algorithms are derived from our new sufficent conditions for topology preservation adapted to some parallel reduction techniques, hence their topological correctness is guaranteed. In addition, the possibility of the maximal reduction for the various algorithms is also examined.

References

[1] R.W. Hall, T.Y. Kong, and A. Rosenfeld. Shrinking binary images, In: T.Y. Kong and A. Rosenfeld (Eds.): *Topological Algorithms for Digital Processing*, Elsevier Science, 1996, 31–98

[2] R.W. Hall. Parallel connectivity-preserving thinning algorithms, In: T.Y. Kong and A. Rosenfeld (Eds.): *Topological Algorithms for Digital Processing*, Elsevier Science, 1996, 145–179

[3] C. Ronse. Minimal test patterns for connectivity preservation in parallel thinning algorithms for binary digital images. *Discrete Applied Mathematics* **21** 1, 1988, 67–79

Circle Covering and its Application for Telecommunication Networks

Endre Palatinus

The circle packing problem has attracted much attention in the last century, and a variant called packings of equal circles in a square receives attention even nowadays. The objective of it is to give the densest packing of a given number of congruent circles with disjoint interiors in a unit square.

However, its dual problem, the circle covering has not been exhaustively studied so far. We aim to find the "sparsest" covering of the unit square with a given number of congruent circles with overlapping interiors allowed. By sparsest we mean the total covering of the square with congruent circles of minimal radii.

The main difficulty of the problem is the uncertainty of our computations caused by the finite precision of computers. To overcome this we have used interval arithmetic to test if a given setting of circles covers the unit square completely. We developed a branch-and-bound–based method for the previously mentioned testing function. We examined our method in a parallel environment, when the phases of our method can be executed concurrently on several CPU cores.

To test the efficiency of our method we also applied it to a telecommunication-related problem: We would like to find an optimal covering of Hungary with TV-stations for terrestrial signal given the positions of the broadcasting antennas. The signals cover a circular area around the broadcasting stations, and our aim is to minimise the sum of these areas, which is proportional to the power fed into the stations. In this case the radii are allowed to be of different size, and to determine the optimal solution we applied a B&B–based reliable optimisation method.

Keywords: Circle covering, interval arithmetics, reliable computing, parallel computing.

Describing Digital Signal Processing Systems in A Functional Reactive Style

Gábor Páli, Zoltán Gilián, Zoltán Horváth, and Anders Persson

Digital signal processing (DSP) often appears as a main component of specialized embedded systems, like radio base stations in the field of telecommunications. In such systems several discrete steps of DSP transforms are composed into applications to perform a complete function, simultaneously handling multiple flows of data with similar chains of processing configured dynamically and individually. Specifications for DSP systems are often written in a style that emphasizes a compositional style, and algorithms are described by themselves and their different configuration-dependent compositions in different documents. It is the responsibility of the application developer to assemble the different parts into a complete application. The majority of this problem consists of designing the interaction of the parts in the processing chain. Properties of the interaction include method of data transfer, data format, and spatial locality.

DSP algorithms or kernels within the system can be already efficiently described in Feldspar [4], a high-level domain-specific language, but currently there is no solution to describe their interaction and orchestration. As an attempt to fill this gap we propose an extension to Feldspar to address these problems. In practice DSP systems are essentially real-time domain-specific embedded systems with strict constraints on both resource usage and execution time. Our recent work, which is being carried out as part of a joint research project of Ericsson Research, Chalmers University of Technology (Gothenburg, Sweden), and Eötvös Loránd University (Budapest, Hungary), focuses on providing language and compiler support for scheduling multiply running instances of connected DSP kernels in different configurations with deadlines and providing resources for them in an embedded DSP environment. We present a potential design of such systems in a domain-specific style which is supported by efficient and reliable compilation techniques as we have learned in case of Feldspar [5] [6] and as others have shown for generic operating systems [3]. Our goal is to propose a design and a prototype implementation for the required language extensions and their compilation to C.

We have been modeling the current design flow at Ericsson in C, and have also been experimenting with using a notation similar to Haskell Arrows [1] in combination with elements of Functional Reactive Programming [2]. This paper presents the results of our experiments supported by case studies and examples in describing parts of radio base stations. The results can be used in future studies on how to design and implement a system-level layer above the current data-flow layer represented by Feldspar.

Acknowledgments

This work is being supported by Ericsson Software Research, Ericsson Business Network Units and SSF (Sweden), the Hungarian National Development Agency (KMOP-2008-1.1.2), Programul Operaţional Sectional Dezvoltarea Resurselor Umane 2007–2013 (POSDRU/6/1.5/S/3-2008, Romania), and the Swedish Research Council (grant no. 621-2009-5638).

References

[1] J. Hughes: Generalising Monads to Arrows, *Science of Computer Programming 37*, pp. 67–111, May 2000.

[2] H. Nilsson, A. Courtney, J. Peterson: Functional Reactive Programming, Continued, *Haskell Workshop 2002*, Pittsburgh, pp. 51–64, 2002.

[3] P-E. Dagand, A. Baumann, T. Roscoe: Filet-o-Fish: practical and dependable domain-specific languages for OS development. *In 5th Workshop on Programming Languages and Operating Systems (PLOS)*, Big Sky, MT, USA, October 2009.

[4] G. Dévai, M. Tejfel, Z. Gera, G. Páli, Z. Horváth: Feldspar – A Functional Embedded Language for Digital Signal Processing, *8th International Conference on Applied Informatics*, Eger, Hungary, January 27 – 30, 2010.

[5] G. Dévai, M. Tejfel, Z. Gera, G. Páli, Gy. Nagy, Z. Horváth, E. Axelsson, M. Sheeran, A. Vajda, B. Lyckegård, A. Persson: Efficient Code Generation from the High-level Domain-specific Language Feldspar for DSPs, *Workshop on Optimizations for DSP and Embedded Systems, in conjunction with IEEE/ACM International Symposium on Code Generation and Optimization (CGO 2010)*, Toronto, Canada, April 24–28, 2010,

[6] E. Axelsson, K. Claessen, G. Dévai, Z. Horváth, K. Keijzer, B. Lyckegård, A. Persson, M. Sheeran, J. Svenningsson, A. Vajda: Feldspar: A Domain Specific Language for Digital Signal Processing algorithms, *Eighth ACM/IEEE International Conference on Formal Methods and Models for Codesign*, Grenoble, France, July 26–28, 2010.

Traffic Flow Prediction in Service Networks

László Pál and Tibor Csendes

The estimation of traffic flow of service networks becomes a more and more important task nowadays. In the past period, a large number of research has been done to develop flow prediction models, which forecasts the future traffic flows ([1, 2]). Service networks comprise here postal, transportational and communication networks. Postal networks are the direct motivation for this talk.

In such a network, the nodes mark the operations (collection, sorting, delivery) of a flow process, while the edges represent the flow directions. Those nodes which contain just outgoing edges, represent the collection places of the network, while the nodes which have only input edges, represent the distribution (delivery) places. With the mechanization of the postal logistic system, the estimation of the daily loads occurring in the deliverers becomes possible. This automatization provides an efficient facility for counting a large part of the letters.

The problem considered in this talk is how to plan the traffic measurement in the network with minimal cost, if we know the cost of the measurements in the nodes. In the case of a given output node, we are looking for those nodes, which influence the traffic of that output node. We want to ensure a preset precision for the output node values in terms of uncertainty intervals. Our aim is to achieve the result with the smallest measurement cost. Similar optimization problems in postal networks are formulated in ([4, 5]) using different objectives like vehicle factors, time limit and frequency, cost, and so on.

An important task of the investigated problem is the network evaluation. Based on the known data in some node, we can evaluate the whole network in order to update the influenced nodes and edges. The network evaluation is made with the help of interval calculation [3]. Hence, the data in the nodes and along the edges are represented by intervals. During the network evaluation the basic interval arithmetic operations are used and the calculated values are propagated form the input nodes to the output nodes.

In this talk we consider some solution algorithms and compare them on generated test networks.

Acknowledgements. This work was supported by the TÁMOP-4.2.2/08/1/2008-0008 project of the Hungarian National Development Agency.

References

[1] M. Cremer and H. Keller. A new class of dynamic methods for the identification of origin-destination flows, *Transportation Research*. 21B (1987), 17–32.

[2] Y. Ding, P.B. Mirchandani, S.A. Nobe. Prediction of network loads based on origin-destination synthesis from observed link volumes. *Transportation Research Record: Journal of the Transportation Research Board*, 1607/1997 (2007), 95–104.

[3] S.M. Rump. *INTLAB – Interval Laboratory.* In: T. Csendes (ed.): Developments in Reliable Computing, Kluwer Academic Publishers, Dordrecht, 1999, 77–104.

[4] Q. Song, C. Zhang, X. Li, and F. Hao. Genetic algorithm based modeling and optimization of the borough postal transportation network, *Decision and Control, 2007 46th IEEE Conference*, New Orleans, LA, 2007, 2850–2855.

[5] A. Syberfeldt, H. Grimm, A. Ng, and M. Andersson. Simulation-based optimization of a complex mail transportation network, *Proceedings of the 40th Conference on Winter Simulation*, Miami, Florida, 2008, 2625–2631.

Interval Based Sensor Network Localization

László Pál and Tibor Csendes

Most wireless sensor network applications require information about the geographical location of each sensor node. An approximated geographical location is needed for acquiring and managing data, geographic routing, geographic hash tables, energy conservation algorithms. Sensing data without knowing the sensor location is meaningless in environmental sensing applications such as animal habitat monitoring, bush fire surveillance, water quality monitoring and precision agriculture. This makes localization capabilities highly desirable in sensor networks.

A large number of research and commercial location systems have been developed over the past two decades. A general survey is found in [2]. Recently, some localization techniques have been proposed to allow estimating node location using information transmitted by a set of nodes with known positions. In this talk we consider distance-based techniques, which use inter-sensor distance or angle measurements in location calculation. There are many approaches for the implementation of the centralized distance-based algorithms. The most important are the following: multidimensional scaling [5], semidefinite programming [1], simulated annealing [3, 4] and genetic algorithm [6].

The localization problem can be formulated as a nonlinear global optimization problem, so we tried to solve it with interval based global optimization methods.

Acknowledgements. This work was supported by the TÁMOP-4.2.2/08/1/2008-0008 project of the Hungarian National Development Agency.

References

[1] P. Biswas, and Y. Ye. Semidefinite programming for ad hoc wireless sensor network localization. In *Proceedings of the 3-rd International Symposium on Information Processing in Sensor Networks*, pages 46–54, Berkeley, CA, USA, 2004.

[2] J. Hightower and G. Borriello. Localization systems for ubiquitous computing. *Computer*, 34:57–66, 2001.

[3] A.A. Kannan, G. Mao, and B. Vucetic. Simulated Annealing based Wireless Sensor Network Localization. *Journal of Computers*, 1:15–22, 2006.

[4] E. Niewiadomska-Szynkiewicz and M. Marks. Optimization Schemes For Wireless Sensor Network Localization. *International Journal of Applied Mathematics and Computer Science*, 19:291–302, 2009.

[5] Y. Shang, W. Ruml, Y. Zhang, and M. Fromherz. Localization from connectivity in sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, 15:961–974.

[6] Q. Zhang, J. Wang, C. Jin, J. Ye, C. Ma, W. Zhang. Genetic Algorithm Based Wireless Sensor Network Localization. In *Proceedings of the 2008 Fourth International Conference on Natural Computation*, pages 608–613, vol. 1, 2008.

Benchmarking the Relationship of Database Management Systems and Application Servers on the Web

Nóra Sterbinszky

E-commerce websites are only prosperous if they are well designed, implemented, and adequate hardware and software configurations are applied. Narrowing down to such software configuration includes selecting the appropriate combination of database management system, application server and operating system.

I have examined five database management systems, such as Apache Derby, Sun MySQL, Oracle Database, PostgreSQL and IBM DB2. They belong to database servers which are nowadays most frequently used. Four application servers were combined with them under test. They are Apache Tomcat, JBoss, Sun Glassfish and Oracle Weblogic. Operating system was the third dimension: Microsoft Windows and a Linux distribution.

A benchmark named TPC-W (Transaction Processing Performance Council Web benchmark) made it possible to compare the alternatives above.
TPC-W is a transactional web benchmark. The workload is performed in a controlled internet commerce environment that simulates the activities of a business-oriented transactional web server. Among numerous TPC-benchmarks, TPC-W is appropriate for benchmarking the cooperation of database and application servers. Its performance metric is the number of web interactions processed per second (WIPS). It offers three different profiles by varying the ratio of browsing (reading processes) to buying (writing processes): browsing (WIPSb), shopping (WIPS) and ordering (WIPSo).
The primary metrics is the WIPS rate which shows the efficiency of the combination of operating system with database and application server.

References

[1] Anderson Supriano, Gustavo M. D. Vieira, Luiz E. Buzato. Evaluation of a read-optimized database for dynamic web applications, `http://www.ic.unicamp.br/ğdvieira/publications/supriano08.pdf`, (2010).

[2] C. Amza, A. Chanda, A. L. Cox, S. Elnikety, R. Gil, K. Rajamani, W. Zwaenepoel, E. Cecchet, J. Marguerite. Specification and implementation of dynamic web site benchmarks, *WWC-5:Proceedings of the IEEE International Workshop on Workload Characterization*, (2002), pages 3–13.

[3] M. Stonebraker, C. Bear, U. Cetintemel, M. Cherniack, T. Ge, N. Hachem, S. Harizopoulos, J. Lifter, J. Rogers, S. Zdonik. One size fits all? Part 2: Benchmarking studies, *CIDR 2007: Proceedings of the Third Biennial Conference on Innovative Data Systems Research*, (2007).

[4] Transaction Processing Performance Council (TPC) Benchmark^{TM} W (Web Commerce) Specification Version 2.0r, `http://www.tpc.org/tpcw/default.asp`, (2010).

How Much is XML Involved in DB Publishing?

Gyula I. Szabó

XML has been intensive investigated lately, with the sentence, that ”XML is (has been) the standard form for data publishing”, especially in data base area [1].

That is, there are assumptions, that the newly published data take mostly the form of XML documents, especially when databases are involved. That is the reason of the heavy investments applied for researching the handling, quering and comprising XML documents [2],[3].

This study would like to check these assumptions, while investigating the documents accessible over the Internet, possible to go under the surface, into the ”deep WEB”. The investigation focuses on the large scientific databases, but the commercial data stored in the ”deep WEB” will be handled also.

The technique of randomly generated IP addresses will be used to reach publicly accessible sites for testing the whole WEB as suggested by Kevin ChenChuan Chang et al. [4].

The random IP addresses (when accessible) can be used for analyzing the files of the addressed site, we would like to check the amount of xml documents among the entity of files present on the given site.

Another aim of the study is finding a simple attribute to be used for declaring an XML-document being a database (one can assume, if the size of the document is “large enough” it can be accepted as database). This hypothese will be also investigated and an acceptable value of size criterion proposed. By counting the documents that are used for storing databases using the proposed minimal size as criterion, the proportion of databases among the XML documents stored on the WEB can also be estimated. We would like to check a few number of known sites of large scientific databases (first of all biochemical and astronomical databases)

We would also like to investigate the rate of ”masked XML” files, the documents, that declares themselves as XML documents, but in fact they are built up as HTML files.

We don’t want to create a new search engine, and the aims of these investigations cannot be fulfilled by using the known search engines, because they try to find a given text in the documents present on the WEB, while we would like to get statistical data over documents with a given structure (or semi-structure).

These investigations can be repeated in the future in order to get a dynamic picture of the growing rate of the number of the XML documents present on the WEB, and also over the growing rate of the size of the databases stored as XML documents..

Acknowledgements

This work was supported by ELTE IK TAMOP

References

[1] Wenfei Fan and Leonid Libkin: On XML Integrity Constraints in the Presence of DTDs, in Journal of the ACM (JACM), Volume 49 , Issue 3, pp 368 - 406, May 2002.

[2] Ray, I. Muller, M.:Using Schemas to Simplify Access Control for XML Documents, in Lecture Notes in Computer Science, 2004, ISSU 3347, pages 363-368

[3] G. Leighton and D. Barbosa. Optimizing XML Compression. In Proceedings of the Sixth International XML Database Symposium (XSym 2009), LNCS 5679, pp. 91-105, 2009.

[4] K. C.-C. Chang, B. He, C. Li, and Z. Zhang. Structured databases on the web: Observations and implications. Report UIUCDCS-R-2003-2321, Dept. of Computer Science, UIUC, Feb. 2003.

An Agent-Based Model Formalization to Support Distributed Execution

Attila Szabó and Richárd O. Legéndi

Agent-Based Modeling (ABM) is becoming a common computational paradigm among simulation techniques [1]. The need of large scale and robust experimental results quickly arise the need of distributed execution, since the problems are inherently computationally hard and an exhaustive search is infeasible compared to the analytical approaches.

Many studies have been conducted in the field of formalization of agent-based models: the number of existing ABM description languages and tools is countless. Since there is no standard protocol for describing simulations so far (however, researchers place a considerable emphasis on establishing it [2]), nearly every software tool that supports ABM development has it's own model description dialect. As their purpose, their level of formalism differs as well, but many of them shares a common drawback: these are somehow bounded for the sequential implementation of the model, thus cannot be used for describing distributed simulations.

There are formalisms, however, that use a high abstraction level. Some model describing protocols follow a semi-formal way, like the ODD (Overview-Design-Details) protocol [2]. It contains three blocks subdivided into seven sections to help researchers improve their models and increase their scientific credibility. Its main goal is to be the first step for establishing a detailed common format of ABM, that is once initiated, will hopefully evolve as it becomes used by a sufficient proportion of modelers. This protocol describes the model mainly verbally, but contains formal components for specific parts, e.g. suggests the using of UML to specify static structure, or mathematical formalism to declare the dynamics of agents (since the interaction patterns are complex and unambiguity may be avoided). Design patterns - static model descriptions - for agent-based modeling are also available [3]. There are absolutely formal model descriptions as well (like [5]), and even programming languages exist where the language tries to replace the formalism [4].

In this paper we introduce a minimal, abstract formalization of ABMs that aggregates the advantages of the existing model definition approaches. A formalization, that is independent from programming languages, but capable to describe the agent-based models' semantics: the agents (such as state variables, initial state, and behavior), the environment of the agents (the "global" state variables), and model dynamics (agent activities and other events). We describe an algorithm that transforms this high-level description into a distributed program, and plan to conduct extended experiments and further researches, using the formalism reported herein.

Acknowledgements

This work was supported by a grant from the Hungarian Government (TÁMOP-4.2. 1/B-09/1/KMR-2010-0003).

References

[1] Axelrod, R.: The Complexity of Cooperation: Agent-Based Models of Competition and Cooperation, *Princeton University Press, Princeton, New Jersey, USA* (1997).

[2] Grimm, V. et al.: A standard protocol for describing individual-based and agent-based models, *Ecological Modelling 198(1-2), pp. 115-126.* (2006).

[3] Gulyás, L.: Understanding Emergent Social Phenomena – Methods, Tools, and Applications for Agent-Based Modeling, *PhD Thesis pp. 66-79., ELTE* (2005).

[4] Legéndi, R.O. et al.: Modeling Autonomous Adaptive Agents with Functional Language for Simulations, *Lecture Notes in Artificial Intelligence, vol. 5816. pp 449-460., Springer-Verlag* (2009).

[5] Scheutz, M., Schermerhorn, P.: Adaptive Algorithms for the Dynamic Distribution and Parallel Execution of Agent-Based Models, *Journal of Parallel and Distributed Computing, 66, pp. 1037-1051.* (2005).

A Knowledge-Based Approach to Raster-Vector Conversion of Large Scale Topographic Maps

Rudolf Szendrei, István Elek, István Fekete, and Mátyás Márton

The scanned paper based maps in raster image format are suitable for humans, but geoinformatics prefer to use the properly converted, vectorized maps. The important topographic maps are already vectorized in most countries by a cumbersome, manual procedure. However, the task of raster-vector conversion of paper based maps will not become obsolete within the next few years. Newly issued maps and the updating of old ones will still require this activity.

In the IRIS project the authors have elaborated the theoretical background of a raster-vector conversion system, and they have developed the prototype of some components of the system. The aim of the development is to automatize the raster-vector conversion as much as possible. This goal puts an emphasis on the knowledge based approach. This article will focus on the automatic recognition and conversion of the three main types of map symbols, to improve the efficiency of the recognition system.

Point-like symbols are small icons each representing a real object (e.g. a monument). The recognition algorithm tries to identify these symbols based on given symbol patterns. Each connected pixel set under a given size limit will be matched against the data base of patterns.

Surface-like symbols cover a region with a solid color, or with a pattern (e.g. lake or scrub). The procedure first determines the smallest repetitive part (kernel) of the texture which can be identified by the algorithm used for point-like symbols.

In order to identify linear symbols (e.g. roads, railroad) both line style and topology must be recognized. To determine topology a graph is created using the end- and fork-points of the road-like graphics.

Currently, the automatic recognition of some kinds of map symbols (e.g. texts) is beyond the scope. Thus, the vectorized coverage generated automatically does not contain all of the elements occurring on the original raster map. Furthermore, the algorithms used for recognition provide the possibility for human expert's intervention in the case of false detection.

An important point in the expertise of human interpreters is, for example, the knowledge of the order of map layers they have been printed in. The inclusion of this knowledge would make the conversion much more intelligent.

Acknowledgements

The research project is supported by the TÁMOP-4.2.1/B-09/1/KMR-2010-003 project. The authors are at the Faculty of Informatics, Eötvös Loránd University of Budapest, and they are members of the university research group in geoinformatics T.E.A.M. (http://team.elte.hu/)

References

[1] S. Baik, et al.: A naive geography analyst system with cognitive support of imagery exploitation. *Lecture Notes in Artificial Intelligence*, number 2974, pages 40-48, 2004.

[2] B. P. Buttenfield, and R. B. McMaster: *Map generalization: Making Rules for knowledge representation*. Longman Scientific & Technical, 1991.

[3] B. Dezső, I. Elek, and Zs. Máriás: Image processing methods in raster-vector conversion of topographic maps. *Proceedings of AIPR 2009 Conference*, pages 83-86, July 2009.

[4] R. Szendrei, I. Fekete, and I. Elek: Texture based recognition of topographic map symbols. *Proceedings of AIPR 2009 Conference*, pages 7-10, July 2009.

Online Infrastructure Dependency Detection and Tracking

István Szombath

IT infrastructures providing vital business services are becoming more and more distributed and heterogeneous. System management has to assure the appropriate quality of services and keep resource usage at a reasonable level. Structural information, especially the dependencies between IT components, is vital to system management. Without the dependency information it is not feasible to determine the impact of IT component faults on business services. Dependency information has foremost importance in adaptive architectures, like dynamic reconfiguration based self healing systems. For instance configuration consolidation (i.e., reallocation of servers) in a virtualized infrastructure (cloud) is only feasible when the dependency information is known. Due to the widespread use of adaptive architectures, gathering information on the structure of the system becomes increasingly important for practical system management.

Current state of the art shows that with passive observation of network communication (e.g., with NetFlow) reconstruction of the IT infrastructure model is feasible. The reconstructed model represents servers and the communication of servers. Important dependencies between servers (so called service dependency, e.g., a dependency between a web server and a database) can be identified from the reconstructed model, for example using the method presented in [1].

The model (e.g., a labeled graph) of the infrastructure can be very complex and it could change rapidly. However even a huge enterprise class IT infrastructure can be described with only a few types of high level service patterns, such as 3 tier architectures, backups, authentication and mailing solutions, etc. Thus the most part of the infrastructure graph are covered by these service patterns [2]. The drawback and challenges of this approach are the computational complexity of pattern matching, and the typical patterns needs to be collected manually.

Our method builds a labeled graph from passive observation of network communication that represents the IT infrastructure and updates it online. A method is also worked out to identify and track the existence of typical patterns of the IT infrastructure, e.g. a 3 tier architecture. Typical service patterns can be set, and the pattern matcher [3] identifies the matches in the model online. This means upon model update new matches may be found or already found matches can became obsolete. A proof of concept is also presented to collect typical service patterns automatically using graph clustering. The engine is capable to evaluate the patterns in real time, even in large scale. The approach is verified using communication logs of a real IT infrastructure. The framework is also capable to propagate the discovered dependency information to an enterprise class system management model repository (IBM Tivoli CCMDB).

References

[1] Andreas Kind, Dieter Gantenbein, Hiroaki Etoh, *Relationship Discovery with NetFlow to Enable Business-Driven IT Management*, 1st IEEE/IFIP Int. Workshop on Business- riven IT Management, 2006.

[2] Thomas Karagiannis,Konstantina Papagiannaki, Michalis Faloutsos, *BLINC: Multilevel Traffic Classification in the Dark*, ACM SIGCOMM Computer Communication Review, vol. 35, no. 4, pp. 229-240, October, 2005.

[3] Ráth, I., Bergmann, G., Ökrös, A., and Varró, D. *Live Model Transformations Driven by Incremental Pattern Matching*. In Proceedings of the 1st international Conference on theory and Practice of Model Transformations (Zurich, Switzerland, July 01 - 02, 2008).

DeepTest the Static Debugger of Java Programs

Zalán Szűgyi, István Forgács, and Zoltán Porkoláb

Program slicing has first been proposed by Mark Weiser [1, 2]. The slicing criterion is a pair $C =< I, V >$, where I is a program point and V is a subset of program variables. The backward static program slice S with respect to slicing criterion C consists of all parts of the program that have direct or indirect effect on the values computed for variables V at I. Later on the definition is extended to forward slicing where startring from the slicing criterion S those statements are selected, which are directly or indirectly depend on S. Program slicing can be used in various ways. It can be used to maintain or get familiar with a large and complex source code written by a third person. It is useful in debugging, to locate the errors more easy. Several applications such as optimization, program alanysis, impact analysis, information flow control are based on program slicing.

Nowadays the program slicing is a progressive research field in computer science. The original definition of program slice is static and backward slice. It means the method applied on source code and determined those statements, that affected the given expression.

There are slicer tools under development for different programming languages. Frama-C is a suite of tools dedicated to the analysis of the source code of software written in C. It contains slicer and dependency analysis tools.

In this paper we present our tool, called DeepTest [4], which is a static debugger based program slicing of Object Oriented applications written in Java programming language. DeepTest performs static slicing and able to compute both forward and backward slices. One of the real use application of slicing is the size of information, i.e. hundreds of lines of codes that is very difficult to be evaluated. DeepTest solves the problem by providing only the direct influences to the user. In this way DeepTest can be considered as an alternitve of traditional debuggers, without executing the code. DeepTest also provides a dynamic call graph tool which is able to mind polymorphical properties of objects when createting call graphs.

References

[1] Weiser, M.: Program slicing. Proceedings of the 5th International Conference on Software Engineering, pages 439–449, IEEE Computer Society Press, March 1981.

[2] Weiser, M.: Program slicing. IEEE Transactions on Software Engineering, Volume 10, Issue 4, pages 352–357, IEEE Computer Society Press, July 1984.

[3] `http://frama-c.com/`

[4] `http://www.deeptest.com/`

Obtaining Geometrical Properties of Binary Images from two Projections Using Neural Networks

Tamás Sámuel Tasi and Péter Balázs

In Binary Tomography Reconstruction the task is to obtain the two-dimensional cross-section images of three-dimensional objects from a small number of projections. While numerous methods exist for solving this problem, all of them presume certain prior knowledge about the binary image to be reconstructed. Such assumed features are often of geometrical type, like connectedness or convexity. The aim of our work is to extract these geometrical properties from the projections themselves, so that they can ease the task of choosing the proper reconstruction method, or setting its parameters appropriately.

For retrieving this additional information from the projections we used the Artificial Neural Network model, which consists of an interconnected group of artificial neurons. We managed to train neural networks to recognize several different features (hv-convexity, 4-connectedness, etc.) with the use of large datasets containing horizontal- and vertical projection values and the desired output for each input pattern. For each classification task we set the optimal parameters of the network by running small tests in advance, and performing a thorough testing with the configurations proved to be the most promising.

Our experiments [1] verified that certain geometrical information of binary images can be acquired from merely the projections. With the use of neural networks we successfully separated hv-convex discrete sets from random binary images; 8-, but not 4-connected discrete sets from hv-convex polyominoes; and we got good results in the classification of hv-convex binary images and almost hv-convex polyominoes for higher resolutions.

References

[1] Gara, M., Tasi, T. S. and Balázs, P. Learning connectedness and convexity of binary images from their projections. *Pure Mathematics and Applications*, 20(1–2):27–48, 2009.

Introducing Database Slicing Method and Context-Free Export/Import

Dávid Tengeri

Nowadays more and more software packages use a database of some kind. This database could be very large, hence it might be very difficult to backup, to migrate or to debug. In such cases it would be useful to find the relevant slice of a database, and backup, migrate or debug only that part.

One way to reduce the size of the backup is to identify the really meaningful part of the data records. This can be called database slicing, which is similar to program slicing ([1], [3]). We used the elements of program slicing to describe the links between the tables of a database. It is easy to select those records from the database that are really needed for making a backup of important data, migrating data from one database to another or finding data-related bugs in the database with our new approach. There are two types of slicing, namely static and dynamic used in program slicing. We redefined these methods in the context of databases. Static slicing works with tables of the database, while dynamic slicing is based on the records of tables. We devised slicing algorithms for these types as well. The slicing algorithms use a dependence graph as their input and produce a slice of database. The algorithms are based on breadth-first search method.

There are a lot of papers which describe how to save your database content, but these methods saving the whole content of the database or they don't define how can we reload the exported data ([2],[4]). The process of context-free export/import enables moving the results of database slicing from one database to another without heavily depending on the IDs of tables. With this method we can select a small piece of the database content and migrate it somewhere else, without losing connections between the records in the slice and duplicating data in the target database.

Owing this, we defined an XML-based language for storing the result of slicing algorithms. With our document we can describe the schema of a database separate from the database system that we are using and we store only the columns of the tables which are necessary to import the data correctly. The schema of the database determines the structure of the exported data.

Later, we implemented the slicing method and the exporting/importing process using the Drupal content management system.

References

[1] David Willmor, Suzanne M. Embury, Jianhua Shao: Program Slicing in the Presence of Database State, 2004, *20th IEEE International Conference on Software Maintenance*

[2] Dongwon Lee, Murali Mani, Frank Chiu, Wesley W. Chu: NeT & CoT: Translating Relational Schemas to XML Schemas using Semantic Constraints, 2002

[3] Frank Tip:A Survey of Program Slicing Techniques, 1995, *Journal of programming languages, Citeseer*

[4] Joseph Fong, San Kuen Cheung:Translating relational schema into XML schema definition with data semantic preservation and XSD graph, 2004, *Information and software technology, 2005, Elsevier*

Economic Interpretation of GARCH Models: an Agent-Based Simulation Study

Balázs Torma and László Gerencsér

GARCH models, developed by R. Engle and T. Bollerslev, are widely used technical models for volatility reproducing a number of features of real stock price processes, among others volatility clustering. However, GARCH parameters do not have any direct economic interpration. One of the objectives of this talk is to present newly discovered possible qualitative relationships between fundamental economic factors and the parameter values of the GARCH model that best fits the data. This is done by developing a detailed, agent-based fundamental model for the market, and considering the associated price process as real data.

In our novel market microstructure model chartists and fundamentalists trade in a stock market. Chartists predict future stock prices by extrapolating current trend. On the other hand, the belief of the fundamentalists about future stock prices is effected by the information they receive. The information arrival process is modelled by a discrete time Hawkes's process, which captures the feedback effects of market news. Assuming a fixed behavior pattern for individual agents of a given type, modulo random choices, the market structure is defined and parameterized by the distribution of wealth among the two groups of agents as described above. Extensive numerical experiments have shown that market structure is a fundamental factor for the coefficients of the best-fitting GARCH(1,1) model. In particular, we established monotonic dependence between the relative weights defining a market structure and the GARCH(1,1) coefficients. However, the exact relationship will depend on the unknown individual behaviors.

On the other hand, the established ability of the technical model to reproduce data generated by the fundamental model motivates the use of GARCH models to detect changes in the market structure. For this purpose we use the MDL (Minimum Description Length) approach to real-time change detection as developed in the works of L. Gerencsér and J. Baikovicius. A stand-alone component of this algorithm is a novel recursive method for estimating GARCH coefficients, developed and analysed by the authors in cooperation with Zs. Orlovits. We have tested our change-point detection algorithm extensively on simulated data, and experienced excellent detection capability. We have also tried our algorithm on real data, and found that alarms on possible abrupt changes in the market structure occur only before the price trends become consistent, up or down, indicating that a real change in the market dynamics has indeed occured.

Comparison of Programmers' Opinion in Change Impact Analysis

Gabriella Tóth

Change impact analysis is generally seen as a very difficult program comprehension problem. One of the reasons of this difficulty is that there is no universal definition for dependency between software artifacts, only algorithms that approximate the dependencies.

In the meantime, different kinds of such algorithms have been developed by researchers. But which algorithm is the most suitable in a specific situation, which one finds the relevant dependencies in the best way? Finding the most relevant dependencies is difficult, and is essentially a creative mental task.

A possible way to answer this important question is to involve programmers, and hear their subjective opinions based on expertise and experience in program comprehension. In this paper, we present such an experiment. We wanted to know what is the difference between not only some well-known algorithms and programmer's opinion, but between programmers' opinions as well, and hence we conducted a case study. This case study was documented earlier when the focus was on static impact analysis algorithms, not on programmers' opinion.

In this work, we report on our experiment conducted with this goal in mind using a compact, easily comprehensible Java experimental software system, simulated program changes, and a group of programmers (developers, computer science student and PhD students) who were asked to perform impact analysis with the help of different tools and on the basis of their programming experience. We applied several well-known algorithms (callgraph, program slicing, static execute after, historical co-change), JRipples[1], a Java framework for change impact analysis embedded in Eclipse development environment, and used BEFRIEND[2] to evaluate the results given by the programmers.

Now the author shows to which algorithms turned out to be the closest the individual programmer's opinion. Furthermore, the programmers are compared to each other according to their qualification, experiment, and the kind of dependencies identified by them.

References

[1] J. Buckner, J. Buchta, M. Petrenko, and V. Rajlich: JRipples: A tool for program comprehension during incremental change. IWPC, 2005.

[2] L. J. Fülöp, P. Hegedűs, and R. Ferenc: BEFRIEND - a Benchmark for Evaluating Reverse Engineering Tools, accepted, to appear. Periodica Polytechnica, 2009.

Building Dependency Graph for Slicing Erlang Programs

Melinda Tóth and István Bozó

Program slicing [1] is the most well-known method used under impact analysis. Different methods are available to perform program slicing (e.g. dataflow equations, information flow relations, dependency graphs), but the most popular from them are based on the dependency graphs of the program to be sliced [2]. These graphs include data and control dependencies of the program.

There are many forms of using program slicing during the software life-cycle. It can be used in debugging, optimization, program analysis, testing or other software maintenance tasks. For example, using program slicing to detect the impact of a change on a certain point on the program, could help to the programmer to find those test cases which could be affected by a program code change.

Our goal is to adopt the existing methods and to develop new algorithms for program slicing of programs written in a dynamically typed functional programming language, Erlang [3]. We describe the algorithms to define Data Dependency Graphs and Control Dependency Graphs for Erlang, and how to build Program/System Dependency Graphs from them. The dependency graphs are usable to reach the mentioned goal and transform the program slicing to a graph reachability problem. We want to calculate the forward slices of the program, especially for those program parts which are changed after a refactoring [4]. Calculating the forward slices could help the programmers to reduce the number of test cases to be rerun after the transformation.

Acknowledgements

This work was supported by TECH_08_A2-SZOMIN08.

References

[1] M. Weiser: Program slices: Formal, psychological, and practical investigations of an automatic program abstraction method. *PhD thesis, University of Michigan, Ann Arbor, MI, 1979.*

[2] S. Horwitz, T. Reps, and D. Binkley: Interprocedural slicing using dependence graphs. *ACM Transactions on Programming Languages and Systems, 12(1):3546, January 1990.*

[3] Erlang Homepage, *https://www.erlang.org*

[4] M. Fowler, K. Beck, J. Brant, W. Opdyke, and D. Roberts: *Refactoring: Improving the Design of Existing Code.* Addison-Wesley, 1999.

Static Type Checking of Model Transformation Programs

Zoltán Ujhelyi

Model-driven development (MDD [1]) is becoming widely accepted in system and software engineering. MDD facilitates the systematic use of models from the very early phase of the design procedure: high-level, visual engineering models are used to capture system requirements and design, followed by the automatic generation of the source code and configuration files of the target application.

A key factor in the successful adoption of MDD is the use of model transformations utilized for various tasks, such as formal model analysis or code generation. However, model transformations in most cases are still written manually in industrial practice as a regular piece of software.

As more and more complex model transformations are developed, ensuring the correctness of the transformation programs becomes increasingly difficult. Nonetheless, detecting errors is required as they can propagate into the developed application, or invalidate the results of formal analysis.

Methods for ensuring correctness of computer programs such as *static analysis* are applicable for transformation programs as well. Static analysis represents a set of techniques for computing different properties of programs without their execution, used extensively in compilers for optimization and also for program verification.

In case of dynamically typed programming languages (such as Javascript) typing errors are one of the most common programming errors. They most often lead to misleading output rather than a runtime exception making them hard to trace. Static type checker tools address these problem by inferring the types of every program variable and validating all their use.

The current paper presents a static type checker component for early detection of typing errors in model transformation programs. The component was implemented for the VIATRA2 [2] model transformation framework (a general Eclipse-based modeling framework developed at BME-DMIS) based on graph transformations (GT) theory, a declarative, rule-based specification paradigm. Complex model transformations are defined by a combination of abstract state machines (ASM) and graph transformation rules. The ASM parts are dynamically typed that necessitates type checking, while GT rules are statically typed that provides information for efficient type inference.

Our approach describes type safety as constraint satisfaction problems: the type system is mapped to special integer sets; and the type information inferrable from the transformation programs are represented as constraints using these sets. Proper error messages are handled by dedicated back-annotation from the constraint domain. For performance considerations the rules and patterns of transformation programs are analyzed separately, and the partial results are described as pre- and postconditions based on the "design by contract" [3] methodology.

References

[1] Object Management Group. Model Driven Architecture — A Technical Perspective, September 2001. (`http://www.omg.org`).

[2] VIATRA2 Model Transformation Framework. An Eclipse GMT Subproject (`http://www.eclipse.org/gmt/VIATRA2/`).

[3] Bertrand Meyer: Object-oriented software construction (2nd ed.). Prentice-Hall, Inc., 1997.

Resampling of Volumetric Data with GPU Generated Distance Field

Viktor Vad

In volume visualization applications it can be necessary to use a reconstruction method, which interpolates between the voxels of the volume. The most convenient strategy is to use a reconstruction filter (cubic B-Spline, for example) on the discreetly sampled data. Another way is to use Approximated Distance Function. This method has two steps. At the first step, 3D approximated distance fields are generated from the 2D slices, at the second step we can interpolate between these distance fields. At the visualization stage, we can get a resampled image, by extracting the zero level-set of the interpolated 3D distance function. The bottleneck of the method is the algorithm which is used to generate the distance fields. There are fast algorithms for this purpose (chamfering, fast marching methods), but these methods lack the desired accuracy. More accurate methods are known, but the usages of them are limited by their computational cost. It is proven by many examples, that the computational time can be highly reduced, if we use the parallel and scalable architecture of the modern graphics processor. But according to our experience, developing an efficient GPU re-implementation can be far from being evident, due to the GPU's special SIMD architecture. Despite that, this work attempts to create a tool to approximate a Distance Function with the NVIDIA CUDA C technology. This tool would benefit of the speed of the GPU, and it would be at least as accurate as the previous fast methods. Then the tool would be used to resample volumetric data, such as CT and MRI scans.

Object Rotation Effects on Binary Tomographic Reconstruction

László Varga, Péter Balázs, and Antal Nagy

The main goal of transmission tomography is to reconstruct the inner structure of objects from their projections. Sometimes, acquiring projections can be of extremely high cost or taking too many of them can damage the object of study. In binary tomography we make the restriction that the object to be reconstructed can only consist of two known materials (usually the parts of a homogeneous object and empty space between them), and sometimes we can also assume that its shape fulfils some special properties as well. With such prior information several algorithms have been developed capable of reconstructing objects from only a few (say, up to 10) projections.

It has already been shown [1, 2, 3] that certain projections can contain more information than others. This fact can be especially crucial in the case of discrete tomography where only a handful of projections are available and choosing appropriate projection angles can yield better reconstruction results.

Experiments on such dependency on the choice of projections have already been presented [3] by comparing the binary tomographic reconstructions of the same objects but with different projection angles. Our research extends the previous work and investigates whether the results still hold in a practical application when the projection data is affected by noise of different characteristics and measurement errors. We have performed experimental tests on a set of software phantoms, comparing their reconstructions from different projection sets corrupted by different levels of random noise.

Our results indicate that the previous results can be extended to the case when the projection data is corrupted by noise and we can discover connections between the reconstructions of objects belonging to the different noise levels. Based on our experiments we discuss a possible application of the results in the field of non-destructive testing as well.

References

[1] A. Nagy, A. Kuba, Reconstruction of binary matrices from fan-beam projections, *Acta Cybernetica*, 17(2) (2005) 359–385.

[2] G. Placidi, M. Alecci, A. Sotgiu, Theory of adaptive acquisition method for image reconstruction from projections and application to EPR imaging, *Journal of Magnetic Resonance*, Series B, (1995) 50–57.

[3] L. Varga, P. Balázs, A. Nagy, Direction-dependency of a binary tomographic reconstruction algorithm, *Lecture Notes in Computer Science* vol. 6026 (2010) 242–253.

Characterization of Semi-CNS Polynomials

Péter Varga

Let $P(x) = c_0 + c_1x + ... + c_{n-1}x^{n-1} \in \mathbb{Z}$ and $D = \{0, 1, ..., c\_0 - 1\}$. The polynomial $P(x)$ is called CNS-polynomial if every coset of the factor ring $\mathbb{Z}[x]/P(x)\mathbb{Z}[x]$ has a member of form

$$\sum_{h=0}^{\infty} d_h x^h, \tag{1}$$

with $d_h \in D, h = 0, 1, \ldots$ and such that only finitely many d_h are non-zero. This concept generalizes the negative-base radix representation of integers. It was introduced and studied in [2]. The characterization of CNS polynomials already for degree three is complicated, as indicated in [3]. It is still unsolved.

Burcsi and Kovács [1] called $P(x)$ a *semi-CNS polynomial* if the finite expansions (1) form an additive semigroup. This is a generalization of the usual radix representations of natural numbers. They were able to prove some sufficient properties for $P(x)$ being a semi-CNS polynomial. Moreover they generalized Brunotte's algorithm for semi-CNS polynomials.

In this talk, which is based on a joint work with A. Pethő we give a complete characterization of cubic semi-CNS polynomials. More precisely, in all those polynomials, which do not satisfy the condition given by Burcsi and Kovács, are not semi-CNS, so all those cubic monic polynomials which has a negative coefficient in addition to the constant one are not semi-CNS. To prove this we present to each polynomial a cycle.

References

[1] Péter Burcsi and Attila Kovács: *Exhaustive search methods for CNS polynomials*, Monatsh. Math. 155, 421-430 (2008).

[2] S. Akiyama, T. Borbély, H. Brunotte, A. Pethő, and J. Thuswaldner: *Generalized radix representations and dynamical systems I*, Acta Math. Hungar., **108 (3)** (2005), 207-238.

[3] S. Akiyama, H. Brunotte, A. Pethő, and J. M. Thuswaldner: *Generalized radix representations and dynamical systems II*, Acta Arith., **121** (2006), 21-61.

Parameter Estimation of a Flow-Measurement Function Used in Digital Angiography

Krisztián Veress

The purpose of angiographical procedures [1], [2] used in cardiovascular interventions is to classify the patient's potential of regeneration after strokes caused by dead blood cells in the main arteria. The main question of a surgeon is whether the affected cardiac muscle is able to regenerate by allowing blood to flow in and out, or not. The flow of blood into heart's capillaries is measured using x-ray radiometry with contrastive fluids [3], [4].

The intensity of the contrastive fluid flowing into the target cardiac muscle area is determined by means of image processing on angiogramms produced during the radiometry, thus resulting a dataset of positive intensity values. Our task was to estimate the parameters of a 5-parameter Gamma function — which is well-verified for such problems in this domain — which provides valueable information to the medical team about the patient's status. The parameter-estimation of the former function is hard given that the raw dataset is heavily polluted with several different noise types.

We propose a solution for eliminating the noise by applying a specially designed moving window Gauss filter. We have successfully verified the proposed smoothing algorithm which lets us to use the Levenberg-Marquardt local search method on the smoothed dataset. The parameter estimation is done in a nonlinear least-squares way. Our results showed that higher precision, lower processing times and faster convergence could be achieved using the smoothed dataset rather than the original one.

Moreover, we designed an algorithm for computing an initial guess for the LM algorithm in order to achieve ultimate precision. The method is based on the functional analysis of the Gamma model, and statistical computations on the original dataset. Feeding the LM with the precomputed initial guess, the residuals decreased apace.

Finally, a third algorithm is proposed for selecting significant points on the smoothed dataset with an interval-based classification method. By selecting 10 to 30 significant points from the average of 200 data points showed that the parameters of the Gamma model coudl be approximated more precisely to the correct values in a biological (not numerical) sense.

Acknowledgements

This work was supported by the Cardiovascular Research Laboratory at the Medical University of Szeged.

References

[1] C.M. Gibson, S. Albert *Angiographic Assessment of Both Epicardial and Myocardial Perfusion*, Circulation, **109**, 3096-3105, (2004)

[2] J. Hung-Chi Yu *3D Digital Subtraction Angiography*

[3] C.M. Gibson, M. Anshelevich, S. Murphy *Impact of injections during diagnostic coronary angiography on coronary patency in the setting of acute myocardial infarction from the TIMI trials*, Am J Cardiol., **86**, 1378-1379, (2000)

[4] C.M. Gibson, A.J. Kirtane, S.A Murphy *Impact of contrast agent type (ionic versus non-ionic) used for coronary angiography on angiographic, electrocardiographic and clinical outcomes following thrombolytic administration in acute MI.*, Catheter Cardiovasc Interv, **53**, 6-11, (2001)

Radiocommunication Testbed for Wireless Sensor Networks

Krisztián Veress

In wireless sensor networks there is a huge need for a test environment or framework where multiple network topologies [1], [2], [3] along with their parameters and applied communication protocols can be easily described and tested under different circumstances. Such a testbed could reveal the effectiveness or the bottleneck of built-up networks and protocols.

Given the problem described above, our purpose was to implement a testbed being capable of performing different types of unit tests focusing on the wireless communication by collecting specially designed statistical indices on it. The framework is written in NesC language under the TinyOS operating system which are the de facto standards for writing embedded applications for wireless sensor networking hardware [4], [5].

Because of the lack of dynamic memory allocation in current TinyOS distributions, we had to overcome the problem of dynamically assigning network topologies and communication schemes to the sensor nodes building up the network. This is done by predefining statically the supported network types and applying configuration procedures realtime.

Our model represents the general networks as a directed graph having sensor nodes as vertices and communication lines as edges. For each edge, the communication scheme can be separately set up. The frameworks supports every message transmission modes (broadcasting, direct addressing and acknowledgements) provided by the TinyOS system along with the Low-Power-Listening [6] feature used mainly in resource-limited applications.

Since WSNs are heavily event-driven, the message transmissions are tied to specific events such as timer triggering, message sending ans reception, or special control messages. During the test runs, simple messages are transmitted having unique payloads that let us to collect statistics about the communication in progress. These are among others: *the count of message sending requests*, *total sent messages*, *resend count*, *sent messages for which acknowledgement has not been received*, *receive count*, *missed message count*, etc. We have established a few equations that must hold between these indices in order to have a method to verify them.

Since statistics are collected per edge, the user has the possibility to evaluate and analyze either a certain communication line (part of a network) or the network in whole.

References

[1] J. Polastre, J. Hill, D. Culler *Versatile Low Power Media Access for Wireless Sensor Networks*, SenSys'04, November 3–5, 2004, Baltimore, Maryland, USA.

[2] A. Arora, P. Dutta, S. Bapat, V. Kulathumani *Wireless Sensor Networks for Habitat Monitoring* WSNA'02, 2002, Atlanta, Georgia, USA

[3] A. Arora, P. Dutta, S. Bapat, V. Kulathumani *A Line in The Sand: A Wireless Sensor Network for target Detection, Classification, and Tracking* Computer Networks Elsevier, (2004), 46, 605–634

[4] S. Madden, et al., *TAG: A Tiny AGgregation Service for Ad-Hoc Sensor Networks* OSDI, 2002

[5] J. Mclurkin, *Algorithms for distributed sensor networks* Masters Thesis for Electrical Engineering at the University of California, Berkeley, 1999.

[6] R. Jurdak, P. Baldi, C.V. Lopes, *Adaptive Low Power Listening for Wireless Sensor Networks*, IEEE Transactions on Mobile Computing, 6(8), (August 2007) 988–1004.

List of Participants

Antal, Bálint: Faculty of Informatics, University of Debrecen, H-4032 Debrecen, Egyetem tér 1., Hungary, E-mail: antal.balint@inf.unideb.hu

Árgilán, Viktor: University of Szeged, JGYTF, Department of Computer Science, H-6725 Szeged, Hungary, E-mail: gilan@jgypk.u-szeged.hu

Balázs, Dávid: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: David.Balazs@stud.u-szeged.hu

Bartók, Tamás: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: tbartok@inf.u-szeged.hu

Bársony, István: Kecskemét College Faculty of Mechanical Engineering and Automation, H-6000 Kecskemét, Izsáki u. 10., Hungary, E-mail: barsony.istvan@gamf.kefo.hu

Bekéné Rácz, Anett: Faculty of Informatics, University of Debrecen, H-4032 Debrecen, Egyetem tér 1., Hungary, E-mail: racz.anett@inf.unideb.hu

Bényász, Gábor: Corvinus University of Budapest, H-1093 Budapest, Fővám tér 8, Hungary, E-mail: gabor.benyasz@stud.uni-corvinus.hu

Bolla, Kálmán: Kecskemét College Faculty of Mechanical Engineering and Automation, H-6000 Kecskemét, Izsáki u. 10., Hungary, E-mail: kalman.bolla@gmail.com

Bozó, István: Department of Programming Languages and Compilers, Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: bozo_i@inf.elte.hu

Bozsik, József: Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: bozsik@inf.elte.hu

Bóta, András: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: bandras@inf.u-szeged.hu

Csendes, Áron: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: csaron@inf.u-szeged.hu

Csernenszky, András: OTP Bank, Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: csernenszkya@gmail.com

Demián, Tamás: Budapest University of Technology and Economics, H-1117 Budapest, Magyar Tudósok krt. 2., Hungary, E-mail: demiantamas@vnet.hu

Dobján, Tibor: College of Dunaújváros, H-2401, Dunaújváros Táncsics M. u. 1/A, Hungary, E-mail: dobjan.tibor@mail.duf.hu

Fischer, Ferenc: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: fischerf@inf.u-szeged.hu

Gara, Mihály: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: gara@inf.u-szeged.hu

Gáti, Attila: University of Óbuda, H-1034, Budapest, Bécsi út 96/b, Hungary, E-mail: matgati@gmail.com

Giachetta, Roberto: Department of Software Technology and Methodology, Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: groberto@inf.elte.hu

Griechisch, Erika: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: griechisch.erika@gmail.com

Hantos, Norbert: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: renslay@gmail.com

Hegyháti, Máté: University of Veszprém, H-8200 Veszprém, Egyetem u. 10., Hungary, E-mail: hegyhati@dcs.uni-pannon.hu

Hodován, Renáta: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: reni@inf.u-szeged.hu

Hullár, Béla: Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: hullarb@elte.hu

Hunor, Jakab: Babes-Bolyai University, Kogalniceanu str. 1. RO-400084 Cluj Napoca, Romania, E-mail: jakabh@cs.ubbcluj.ro

Kalauz, Károly: University of Veszprém, H-8200 Veszprém, Egyetem u. 10., Hungary, E-mail: kalauz@dcs.uni-pannon.hu

Kardos, Péter: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: pkardos@inf.u-szeged.hu

Kasza, Miklós: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: kaszam@inf.u-szeged.hu

Kunkli, Roland: Faculty of Informatics, University of Debrecen, H-4032 Debrecen, Egyetem tér 1., Hungary, E-mail: kunkli.roland@inf.unideb.hu

Lázár, István: Faculty of Informatics, University of Debrecen, H-4032 Debrecen, Egyetem tér 1., Hungary, E-mail: lazar.istvan@inf.unideb.hu

Legéndi, Richárd Olivér: Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: legendi@inf.elte.hu

Lengyel, Zsolt: Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: lengyel@caesar.elte.hu

Lócsi, Levente: Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: locsi@inf.elte.hu

Manouchehr, Shaghaghi K.: Faculty of Engineering Sciences, Széchenyi István University, H-9026 Győr, Egyetem tér 1, Hungary, E-mail: manouch_28@hotmail.com

Máriás, Zsigmond: Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: zmarias@inf.elte.hu

Mosavi, Amir: Faculty of Informatics, University of Debrecen, H-4032 Debrecen, Egyetem tér 1., Hungary, E-mail: a.mosavi@math.unideb.hu

Németh, Gábor: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: gnemeth@inf.u-szeged.hu

Palatinus, Endre: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: palatinuse@gmail.com

Pál, László: Sapientia University, 530170 Piata Libertatii, Nr. 1, Miercurea-Ciuc, Romania, E-mail: pallaszlo@sapientia.siculorum.ro

Páli, Gábor János: Department of Programming Languages and Compilers, Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: pgj@elte.hu

Pleva, Péter: Faculty of Informatics, University of Debrecen, H-4032 Debrecen, Egyetem tér 1., Hungary, E-mail: peterpleva@gmail.com

Sterbinszky, Nóra: Faculty of Informatics, University of Debrecen, H-4032 Debrecen, Egyetem tér 1., Hungary, E-mail: snorav@gmail.com

Szabó, Attila: Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: aszabo@aitia.ai

Szabó, Gyula István: Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: gyula@szaboo.de

Szendrei, Rudolf: Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: swap@inf.elte.hu

Szombath, István: Budapest University of Technology and Economics, H-1117 Budapest, Magyar Tudósok krt. 2., Hungary, E-mail: szombath@mit.bme.hu

Szűgyi, Zalán: Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: lupin@ludens.elte.hu

Tasi, Tamás Sámuel: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: h494691@stud.u-szeged.hu

Tengeri, Dávid: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: dtengeri@inf.u-szeged.hu

Tóth, Gabriella: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: gtoth@inf.u-szeged.hu

Tóth, Melinda: Department of Programming Languages and Compilers, Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: toth_m@inf.elte.hu

Tóth, Vivien: Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: vivi.toth@gmail.com

Ujhelyi, Zoltán: Budapest University of Technology and Economics, H-1117 Budapest, Magyar Tudósok krt. 2., Hungary, E-mail: ujhelyiz@mit.bme.hu

Vad, Viktor András: Department of Programming Languages and Compilers, Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: vad@elte.hu

Valasek, Gábor: Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: valasek@inf.elte.hu

Varga, László: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: vargalg@inf.u-szeged.hu

Varga, Péter: Faculty of Informatics, University of Debrecen, H-4032 Debrecen, Egyetem tér 1., Hungary, E-mail: vapeti@gmail.com

Veress, Krisztián: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: veresskrisztian@gmail.com

Notes

Notes