GENERIC DNA ENCODING DESIGN SCHEME TO SOLVE COMBINATORIAL PROBLEMS

ROFILDE HASUDUNGAN

Thesis submitted in fulfillment of requirements
for the award of the degree of
Master of Computer Science

Faculty of Computer System and Software Engineering
UNIVERSITI MALAYSIA PAHANG

December 2015

TABLE OF CONTENT

CHAPTER 4 RESULT AND DISCUSSION

CHAPTER 5 CONCLUSIONS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic Acid
A	Adenine
C	Cytosine
G	Guanine
T	Thymine
TSP	Traveling Salesman Problem
SRCP	Scheduling in Robotic Cell
FMS	Flexible Manufacturing System
VCP	Vertex Coloring Problem
DCLP	Distribution Center Location Problem
NP	Nondeterministic Polynomial
HPP	Hamiltonian Path Problem
SAT	Satisfiability or SATISFIABILITY
ssDNA	Single Strand DNA

GENERIC DNA ENCODING DESIGN SCHEME TO SOLVE COMBINATORIAL PROBLEMS

ROFILDE HASUDUNGAN

Thesis submitted in fulfillment of requirements
for the award of the degree of
Master of Computer Science

Faculty of Computer System and Software Engineering
UNIVERSITI MALAYSIA PAHANG

December 2015

ABSTRACT

Combinatorial problems arise in many areas of computer science and application domains. It involves finding groupings, ordering, or assignment of a discrete, finite set of objects that satisfy given conditions. The complexity of combinatorial problems is classified as NP meaning that algorithms are yet to exist to efficiently solve the problem. However, DNA computing can solve the problem in linear time since the parallel processing power of DNA computing is able to generate a solution in a single process. DNA encoding is the first important step in DNA computing phases. Currently, data encoding in DNA computing is tightly coupled with an algorithm that solves an instance of the problem. Solving another problem requires developing specific encoding and computations anew to prove DNA encoding and form the algorithm which is costly. This study proposes a generic DNA encoding schema capable of representing different combinatorial problems. To render the generic encoding scheme capable of solving the different problems, we introduce graph modelling to describe all possible solutions for the problem, where the parameters are converted into vertices and edges before encoding it into DNA sequences. From graph modelling, we construct the encoding scheme consisting of three parts: (1) vertex that links to another vertex, (2) edges that contain information and (3) vertex that links to another vertex. To prove the concept, we employ four different combinatorial problems: Traveling Salesman Problem, Distribution Centre Location Problem, Scheduling Robotic Cell Problem, and Vertex Colouring Problem. Computer simulations show that the proposed generic encoding can generate the desired solution and biological operations could produce solutions for each problem. This approach was applied successfully to solve four hard combinatorial problems. Using this encoding scheme would enable researchers to solve other hard problems whilst also improving the algorithm.

ABSTRAK

Masalah kombinasi timbul dalam banyak bidang sains komputer dan pemakaian lainnya. Ia melibatkan mencari kumpulan, susunan, atau penyerahan hak diskret, set objek terhingga yang yang sesuai dengan syarat yang diberikan. Kerumitan masalah kombinasi dikelaskan sebagai NP dimana tidak ada algoritma boleh menyelesaikannya dengan cekap. Namun, dengan menggunakan DNA computing boleh menyelesaikan dalam masa linear sebab ianya mempunyai kuasa yang mampu penyelesaian dalam pemprosesan tunggal. Pengekodan DNA adalah fase yang pertama dan satu langkah yang penting pada DNA computing. Pada masa ini, pengekodan data dalam DNA computing sangat berkaitan erat dengan algoritma yang menyelesaikan satu contoh masalah dan menyelesaikan masalah lain harus merangka pengekodan tertentu dan pengiraan menjalankan dari awal untuk membuktikan pengekodan DNA dan algoritma yang mana ianya mahal. Kajian ini mencadangkan skema pengekodan DNA generik yang membolehkan untuk mewakili masalah kombinasi yang berbeza. Untuk membuat skim pengekodan generik boleh digunakan untuk menyelesaikan masalah yang berbeza, kami memperkenalkan model graf untuk menjelaskan semua kemungkinan penyelesaian bagi masalah ini, di mana parameter diubah menjadi vertices dan edges sebelum pengekodan kepada jujukan DNA. Dari pemodelan Graf, kita buat skim pengekodan yang dibina graf, bahawa skim pengekodan yang terdiri daripada tiga bahagian: (1) vertex yang mengandungi pautan ke vertex yang lain, (2) edge yang mengandungi maklumat dan (3) vertex yang mengandungi pautan ke vertex lain. Untuk membukti konsep, kita bekerja empat masalah kombinasi yang berbeza: *Traveling Salesman Problem*, *Distribution Center Location Problem*, *Scheduling in Robotic Cell*, dan *Vertex Coloring Problem*. Menggunakan simulasi komputer menunjukkan bahawa pengekodan generik yang dicadangkan boleh menjana penyelesaian yang dikehendaki dan melalui operasi biologi menghasilkan penyelesaian bagi setiap masalah. Pendekatan ini berjaya digunakan untuk menyelesaikan empat masalah kombinasi yang dianggap masalah yang rumit dan dengan menggunakan skim pengekodan ini membolehkan penyelidik untuk menyelesaikan satu lagi masalah yang sukar serta meningkatkan algoritma.

REFERENCES

Adleman, L. M., 1994. Molecular computational of solutions to combinatorial problem. *Journal of Science,* 226(5187), pp. 1021-1024.

Albayrak, M. & Allahverdi, N., 2011. Development a new mutation operator to solve the Traveling Salesman Problem by aid of Genetic Algorithms. *Expert Systems with Applications,* Volume 38, pp. 1313--1320.

Amos, M., 2004. *Theoretical and experimental DNA Computation.* s.l.:Springer.

Bakar, R. A. & Watada, J., 2007. A bio-soft computing approach to re-arrange a flexible manufacturing robot. In: Apolini & Bruno, eds. *Knowledge-Based Intelligent Information and Engineering Systems.* Vietri sul Mare: Knowledge-Based Intelligent Information and Engineering Systems, pp. 308-315.

Bakar, R. A. & Watada, J., 2008. A biologically inspired computing approach to solve cluster-based determination of logistic problem. *International Journal of Biomedical Soft Computing and Human Science,* Volume 13, No. 2, pp. 59-66.

Bakar, R. A., Watada, J. & Pedrycz, W., 2008. DNA approach to solve clustering problem based on mutual order. *Biosystems,* 91(1), pp. 1-12.

Batur, G. D., Karasan, O. E. & Akturk, M. S., 2012. Multiple part-type scheduling in flexible robotic cells. *International Journal of Production Economics,* Volume 135, pp. 726--740.

Bhattacharya, R. & Bandyopadhyay, S., 2010. Solving conflicting bi-objective facility location problem by NSGA II evolutionary algorithm. *The International Journal of Advanced Manufacturing Technology,* Volume 51, pp. 397-414.

Brenneman, A. & Condon, A., 2002. Strand design for biomolecular computation. *Theoretical Computer Science,* Volume 287, pp. 39--58.

Burnwal, S. & Deb, S., 2013. Scheduling optimization of flexible manufacturing system using cuckoo search-based approach. *The International Journal of Advanced Manufacturing Technology,* Volume 64, pp. 951--959.

Chen, S. & Chien, C., 2011. Solving the traveling salesman problem based on the genetic simulated annealing ant colony system with particle swarm optimization techniques. *Expert Systems with Applications,* Volume 38, pp. 14439-14450.

Church, G. M., Gao, Y. & Kosuri, S., 2012. Next-Generation Digital Information Storage in DNA. *Journal of Science,* 337(6102), p. 1628.

Daley, M. J. & Kari, L., 2002. DNA Computing: Models and Implementatios. *Comment on Theretical Biology,* Volume 7, pp. 177-198.

Dawande, M., Geismar, H. N., Pinedo, M. & Sriskandarajah, C., 2009. Throughput optimization in dual-gripper interval robotic cells. *IIE Transactions,* Volume 42, pp. 1--15.

Formanowicz, P., 2005. DNA Computing. *Computational Methods in Science and Technology,* 11(1), pp. 11-20.

Furini, F. & Malaguti, E., 2012. Exact weighted vertex coloring via branch-and-price. *Discrete Optimization,* 9(2), pp. 130-136.

Geismar, H. N. & Pinedo, M., 2010. Robotic cells with stochastic processing times. *IIE Transactions,* Volume 42, pp. 897--914.

Gnanavel Babu, A. et al., 2010. Scheduling of machines and automated guided vehicles in FMS using differential evolution. *International Journal of Production Research,* Volume 48, pp. 4683-4699.

Goos, M., Lempiainen, T., Czeizler, E. & Orponen, P., 2014. Search methods for tile sets in patterned DNA self-assembly. *Journal of Computer and System Sciences,* Volume 80, pp. 297-319.

Govor\vcin, J., Gvozdenovi\c, N. & Povh, J., 2013. New heuristics for the vertex coloring problem based on semidefinite programming. *Central European Journal of Operations Research,* Volume 21, pp. 13--25.

Gualandi, S. & Malucelli, F., 2012. A simple branching scheme for vertex coloring problems. *Discrete Applied Mathematics,* Volume 160, pp. 192-196.

Han, A. & Zhu, D., 2006. *A new DNA encoding method for traveling salesman problem.* Kunming, Springerx, pp. 328-335.

Han, A. & Zhu, D., 2008. DNA Encoding Methods in the Field of DNA Computing. In: *Computational Intelligence in Bioinformatics.* s.l.:Springer Berlin Heidelberg, pp. 293-322.

Hartamanis, J., 1995. On the weight of computations. *Bulletin of the European Assosiation for Theoritical Computer Science,* Volume 55, pp. 136-138.

Ibrahim, Z., Kurniawan, T. B., Khalid, M. & Ono, O., 2007. *A DNA Sequence Design for Direct-Proportional Length-Based DNA Computing using DNASequenceGenerator.* Kumamoto, IEEE, pp. 364--364.

Ibrahim, Z., Tsuboi, Y. & Ono, O., 2004. Direct-proportional length-based DNA computing for shortest path problem. *International Journal of Computer Science and Applications,* Volume I, pp. 46-60.

Ibrahim, Z., Tsuboi, Y., Ono, O. & Khalid, M., 2006. Hybrid concentration-controlled direct-proportional length-based DNA computing for numerical optimization of the shortest path problem. *Biologically Inspired Approaches to Advanced Information Technology*, Volume 3853, pp. 206--221.

Ignatova, Z., MartÃnez-PÃ©rez, I. & Zimmermann, K.-H., 2008. *DNA computing models.* s.l.:Springer.

Ito, Y. & Fukusaki, E., 2004. DNA as a 'Nanomaterial'. *Journal of Molecular Catalysis B: Enzymatic* , Volume 28 , pp. 155-166.

Jeng, D. J.-F., Kim, I. & Watada, J., 2008. Bio-soft computing with fixed-length DNA to a group control optimization problem. *Soft Computer,* Volume 12, pp. 223-228.

Karaboga, D. & Gorkemli, B., 2011. *A combinatorial artificial bee colony algorithm for traveling salesman problem.* Istanbul, IEEE, pp. 50-53.

Kari, L. & Landweber, L. F., 1999. Computing with DNA. *Bioinformatics Methods and Protocols*, Volume 132, pp. 413--430.

Kari, L. & Mahalingam, K., 2010. DNA computing: a research snapshot. In: J. M. Atallah & M. Blanton, eds. *Algorithms and theory of computation handbook.* s.l.:Chapman & Hall/CRC, pp. 31--31.

Kari, L. & Mahalingam, K., 2010. Watson-Crick palindromes in DNA computing. *Natural Computing,* Volume 9, pp. 297--316.

Kats, V. & Levner, E., 2011. A faster algorithm for 2-cyclic robotic scheduling with fixed robot route and interval processing times. *European Journal of Operation Research,* Volume 209, pp. 51-56.

Kostal, P. & Velisek, K., 2011. Flexible Manufacturing System. *Word Academy of Science, Engineering and Technology,* Volume 77, pp. 825-829.

Kumar, M. S., Janardhana, R. & Rao, C., 2011. Simultaneous scheduling of machines and vehicles in an FMS environment with alternative routing. *The International Journal of Advanced Manufacturing Technology,* Volume 53, pp. 339--351.

Kuo, M. -S., 2011. Optimal location selection for an ternational distribution center by using a new hybrid method. *Expert Systems with Applications,* Volume 38, pp. 7208-7221.

Lee, J. Y. et al., 2003. Temperature gradient-based DNA computing for graph problem with weighted edges. *DNA 8, LNCS 2568,* Volume 2568, pp. 73-84.

Lee, J. Y., Shin, S. -Y., Park, T. H. & Zhang, B. -T., 2004. Solving traveling salesman problems with DNA molecules encoding numerical values. *BioSystem,* Volume 78, pp. 39-47.

Lee, Y. H. & Kwon, S. G., 2010. The hybrid planning algorithm for the distribution center operation using tabu search and decomposed optimization. *Expert Systems with Applications,* Volume 37, pp. 3094-3103.

Lipton, R. J., 1995. Using DNA to solve NP-Complete problems. *Science,* Volume 268, pp. 542-545.

Liu, S., Chan, F. T. & Chung, S., 2011. A study of distribution center location based on the rough sets and interactive multi-objective fuzzy decision theory. *Robotics and Computer-Integrated Manufacturing,* Volume 27, pp. 426--433.

Liu, X. & Tang, H., 2009. *A migrating DNA computing technique for grid clustering.* Taipei, IEEE, pp. 249-252.

Li, Y., 2007. *Traveling salesman problem based on DNA computing.* Haikou, IEEE, pp. 28-34.

Li, Y., Liu, X. & Chen, Y., 2011. Selection of logistics center location useing Axiomatic Fuzzy Set and TOPSIS methodology in logistics management. *Expert Systems with Applications,* Volume 38, pp. 7901-7908.

Maazallahi, R., Niknafs, A. & Arabkhedri, P., 2013. A polynomial-time DNA computing solution for the N-Queens problem. *Procedia-Social and Behavioral Sciences,* Volume 83, pp. 622-628.

Manrique, D., Rodriguez-Paton, A. & Sosik, P., 2011. On the scability of biocomputing algorithms: The case of the maximum clique problem. *Theoritical Computer Science,* Volume 412, pp. 7075-7086.

Marinakis, Y., Marinaki, M. & Dounias, G., 2011. Honey bees mating optimization algorithm for the Euclidean traveling salesman problem. *Information Sciences,* Volume 181, pp. 4684--4698.

Moore, G. E., 1965. Cramming more components onto integrated circuits. *Electronics,* Volume 38, p. 114 ff.

Morcom, A. M., Good, C. D., Frackowiak, R. S. & Rugg, M. D., 2003. Age effects on the neural correlates of successful memory encoding. *Brain,* Volume 126, pp. 213--229.

Muhammad, M. S., Ibrahim, Z. & Khalid, M., 2005. *Direct-Proportional Length-Based DNA Computing Implementation for Elevator Scheduling Problem.* Melborne, IEEE, pp. 1-5.

Mukherjee, S., Ganguly, D., Bhattacharya, S. & Mukherjee, P., 2011. A Cognitive Study on DNA Based Computation. *International Journal of Recent Trends in Engineering,* Volume I, pp. 51-56.

Narayanan, A., 1997. Representing arc labels in DNA algorithms. *Exeter: University of Exeter.*

Narayanan, A. & Zorbalas, S., 1998. *DNA algorithms for computing shortest paths.* Madison, IEEE, pp. 718-723.

Narsingh, D., 2004. *Graph Theory with Applications to Enginnering with Computer Science.* New Delhi: PHI Learning Pvt. Ltd..

Niu, Y., Zhang, X. & Cui, G., 2011. *Solving Maximum Clique Problems with Microfluidic DNA Computer.* Hangzhou, IEEE, pp. 224-227.

Ouyang, Q., Kaplan, P. D., Liu, S. & Libchaber, A., 1997. DNA solution of the maximal clique problem. *Science,* Volume 278, pp. 446-449 .

Paun, G., Rozenberg, G. & Salomaa, A., 1998. *DNA Computing New Computing Paradigms.* Leiden: Springer.

Renna, P., 2011. Multi-agent based scheduling in manufacturing cells in a dynamic environment. *International Journal of Production Research,* Volume 49, pp. 1285--1301.

Ren, Y., Xing, T., Chen, X. & Chai, X., 2011. Fuzzy comprehensive evaluation of location plan of city distribution center. *Energy Procedia,* Volume 13, pp. 479-484.

Rezzazi, M. & Roayaei, M., 2011. Using sticker model of DNA computing to solve domatic partition, kernel and induced path problems. *Information Sciences,* Volume 181, pp. 3581-3600.

Sadeg, S., Drias, H., Aid, H. & Mazouz, S., 2010. *DNA based algorithms for solving both MAX-SAT and MAX-W-SAT problems.* Changsha, IEEE, pp. 195-199.

Sanches, C. A. A. & Soma, N. Y., 2014. A computational DNA solution approach for the quadratic diophantine equation. *Applied Mathematics and Computation,* Volume 238, pp. 436-443.

Shi, N.-Y. & Chu, C.-P., 2010. A molecular solution to the hitting-set problem in DNA-based supercomputing. *Information Sciences,* Volume 180, pp. 1010-1019.

Shin, S. -Y., Zhang, B. -T. & Jun, S. -S., 1999. *Solving traveling salesman problem using molecular programming.* Washington, DC, IEEE, pp. 994-1000.

Smith, M. L., Ramesh, R., Dudek, R. A. & Blair, E. L., 1986. *Characteristics of US flexible manufacturing systems-a survey.* Amsterdam, Elsevier Science Publishers, pp. 477-486.

Souza, A., 2011. *Combinatorial Algorithms.* Berlin: Humbolt University.

Sun, H., Gao, Z. & Wu, J., 2007. A bi-level programming model and solution algorithm for the location of logistics distribution centers. *Applied Mathematical Modelling,* 32(4), pp. 610-616.

Taghipour, H., Taghipour , A., Rezaei, M. & Esmaili, H. A., 2012. Solving the independent set problem by sticker based DNA computers. *American Journal of Molecular Biology,* Volume 2, p. 153.

Tang, Q. & Xie, F., 2009. *Distribution center location optimization by genetic algorithm.* Tianjin, IEEE, pp. 196-199.

Torkestani, J. A., 2013. A New Approach to the Vertex Coloring Problem. *Cybernetics and Systems,* Volume 44, pp. 444-466.

Torkestani, J. A. & Meybodi, M. R., 2011. A cellular learning automata-based algorithm for solving the vertex coloring problem. *Expert Systems with Applications,* Volume 38, pp. 9237--9247.

Torkestani, J. A. & Meybodi, M. R., 2012. A New Vertex Coloring Problem based on Variable Action-Set Learning Automata. *Computing & Informatics,* Volume 29, p. 447–466.

Wang , Z., Huang, D., Meng, H. & Tang, C., 2013. A new fast algorithm for solving the minimum spanning tree problem based on DNA molecules computation. *BioSystem,* Volume 114, pp. 1-7.

Wang, Q. -H. et al., 2009. *DNA algorithm based on incomplete molecule commixed encoding for the shorthest path problem.* Nanjing, IEEE, pp. 3641-3644.

Xikui, L. & Yan, L., 2009. *DNA computing for Traveling Salesman problem.* Beijing, IEEE, pp. 1-4.

Xue, J. & Liu, X., 2011. *Applying DNA computation to clustering in graph.* Deng Feng, IEEE, pp. 986-989.

Xu, J. et al., 2011. An unenumerative DNA computing model for vertex coloring problem. *IEEE Transactions on NanoBioscience,* Volume 10, pp. 94-98.

Xu, J. et al., 2010. *A parallel type of DNA computing model for graph vertex coloring problem.* Changsha, IEEE, pp. 231-235.

Xu, Z. et al., 2010. *A novel and fast parallel solution for the vertex cover problem on DNA-based supercomputing.* Changsha, IEEE, pp. 190-194.

Yamamoto, M. et al., 2002. Solutions of shortest path problems by concentration control. In: N. Jonoska & N. C. Seeman, eds. *DNA Computing.* s.l.:Springer, pp. 203--212.

Yamamoto, M. et al., 2002. Solutions of shorthest path problems by concentration control. *DNA Computing: 7th International Workshop on DNA-Based Computers,* Volume 2340, pp. 203-212.

Yang, J. & Xu, J., 2012. *A molecular computing model for maximal clique problem using circular DNA length growth.* Harbin, IEEE, pp. 558--562.

Yang, J. & Xu, J., 2012. *A molecular computing model for maximal clique problem using circular DNA length growth.* Harbin, IEEE, pp. 558 - 562.

Yibo, D., Rui, W. & Wei, S., 2011. *Location selection formula of logistics distribution center research based on immune genetic algorithm.* Chengdu, s.n., pp. 2382-2387.

Zahrouni, W. & Kamoun, H., 2012. Sequencing and scheduling in a three-machine robotic cell. *International Journal of Production Research,* Volume 50, pp. 2823--2835.

Zhang, B.-T. & Shin, S.-Y., 1998. Molecular Algorithms for E cient and Reliable DNA Computing. *Genetic Programming,* Volume 98, pp. 735-742.

Zhang, H. & Liu, X., 2011. A CLIQUE algorithm using DNA computing techniques based on closed-circle DNA sequences. *BioSystem,* 105(1), pp. 73-82.

Zhang, H., Yu, X. & Zhai, Y., 2014. *A DNA model for solving the hierarchical clustering problem.* Shenzhen, IEEE, pp. 531 - 534.

Zhang, X., Niu, Y., Li, F. & Gan, Z., 2011. *Solving minimum vertex cover problems with microfluidic DNA computer.* Shenzhen, IEEE, pp. 509 - 513.

Zhang, Z., Fan, R., Wang, Y. & Cui, G., 2012. *3D DNA Self-Assembly for the minimum clique problem.* Beijing, IEEE, pp. 438-443.

Zheng, Z. & Zheng , Z., 2010. *A logistics distribution center location method based on improved genetic algorithm.* Shanghai, IEEE, pp. 123-126.

Zhinxiang , Y., Bosheng, S., Cheng , Z. & Cheng, H., 2010. *Molecular beacon-based DNA computing model for maximum independent set problem.* Changsha, IEEE, pp. 732-735.

Zindros , D., n.d. *A Gentle Introduction to Algorithm Complexity Analysis.* [Online] Available at: http://discrete.gr/complexity/ [Accessed 7 September 2015].

Zoraida, B. S. E., 2011. *DNA algorithm employing temperature gradient for chinese postman problem.* Coimbatore, IEEE, pp. 1-4.

APPENDIX A1

LIST OF PUBLICATION

1. Rofilde Hasudungan, Rohani Abu Bakar, Rozlina Mohamed, 2013. *DNA Computing and Its Applications on NP-Còmpleteness Problem.* International Conference on Software Engineering & Computer System.
2. Rofilde Hasudungan, Rohani Abu Bakar, 2013. *DNA Computing for Solving Distribution Center Location Problem.* IEEE International Conference on Control System Computing and Engineering.
3. Rofilde Hasudungan, Rohani Abu Bakar, 2013. *DNA Computing to Solve Vertex Coloring Problem.*
4. Rofilde Hasudungan, Rohani Abu Bakar. 2015. *Generic DNA Computing Design Schema to Solve Combinatorial Problem*