Available online at www.elixirpublishers.com (Elixir International Journal)

Discrete Mathematics

Elixir Dis. Math. 32 (2011) 2003-2006

ISSN: 2229-712X

# Fast parallel DNA algorithm based on adleman-lipton model: the independent dominating set problem

Kavitha Joseph

Department of Mathematics, CMR Institute of Technology, Bangalore, India

**ARTICLE INFO**

**Article history:**

Received: 29 January 2011;

Received in revised form:

22 February 2011;

Accepted: 1 March 2011;

**Keywords**

DNA computing,
Independent dominating set problem,
DNA based algorithm,
NPcomplete,
Biological operations,
Time complexity,
Adleman-Lipton model.

**ABSTRACT**

The independent dominating set problem is a classical optimization problem and has been shown to be NP-Complete. This study finds a molecular computing model to solve the independent dominating set problem, based on Adleman-Lipton model. It proves how to apply stickers in the sticker based model to construct the DNA solution space of the independent dominating set problem and how to apply DNA operations in the Adleman-Lipton model to solve that problem from the solution space of stickers. The time complexity of the proposed computational model is $O(n + 2m)$ and to verify this model, a small independent dominating set problem was solved. This proves the capacity of molecular computing for solving the complex independent dominating set problem.



## Introduction

The dominating and independent dominating set problem arises in various important applications, including in a communication network. These are important structures, and many optimization approaches rely on these. In clustering schemes, independent sets result in clusterheads that have local control of their cluster without interference. Additionally, adominating independent set based clustering scheme ensures that the entire network is covered.

For example, especially in energy-efficient wireless computing, clustering allows for some wireless nodes to perform fewer tasks by delegating them to their respective clusterhead. On the other hand, the tasks of these clusterheads then result in additional energy consumption. Choosing a few clusterheads according to a minimum independent dominating set, result in energy savings for the network [17]. Several approaches have been developed to solve the independent dominating set problem. This study introduces an alternative DNA-based computing approach to solve the independent dominating set problem. DNA-based computing is an exciting contemporary multidisciplinary research area at the intersection of mathematics, computer science and molecular biology. Through advances in molecular biology [1, 15], it is now possible to produce 1018 or more DNA strands in a tube. Those 1018 or more DNA strands can also be applied for representing 1018 or more bits of information. Biological operations can be used to simultaneously operate 1018 or more bits of information. Or we can say that 1018 or more data processors can be executed in parallel. The power of parallel, high density computation by molecules in solution allows DNA computers to solve hard computational problems such as NP-complete problem in polynomial increasing time.

Adleman [1], who solved a Hamiltonian path problem (HPP) for a directed graph with seven nodes, established the efficacy of the molecules in solution to solve computational problems. Lipton [13] solved a satisfiability (SAT) problem to demonstrate the advantages of exploiting the massive parallelism that is inherent in DNA-based computing. In recent years methods for solving several well known NP-complete problems [5-11] have been proposed using DNA-based computation.

This study finds a molecular solution to the independent dominating set problem in three steps. In the very first step, DNA strands are defined to represent the vertices of the graph with adjacency relation. In second step, the DNA computational model is adopted to construct the algorithm; and in the third step, the read operation is used to figure out the solution of the independent dominating set problem. The rest of this study is organized as follows: In Section 2, the Adleman-Lipton model is introduced. Section 3 presents the DNA based algorithms to generate the solution space for the independent dominating set problem and to solve the independent dominating set problem. Section 4 explains the implementation of the algorithm. Section 5 provides the time complexity of the algorithm. Section 6 checks the correctness of the algorithm. Discussions and conclusions are drawn in Section 7.

## DNA based super computing

The model used in this paper employs only the mature DNA biological operations. This model took the biological operations in the Adleman-Lipton model [1] and the solution space of stickers [2] in the sticker based model.

## The Adleman-Lipton model

A DNA (Deoxyribo Nucleic Acid) is a polymer, which is strung together from monomers called deoxyribonucleotides [15].

Tele:
E-mail addresses: kavijoseph_cmrit@rediffmail.com

Distinct nucleotides are detected only with their bases. Those bases are respectively, abbreviated as A, G, C and T. Two strands of DNA can form a double strand under appropriate conditions, if the respective bases are the Watson-Crick complements of each other-

A matches T and C matches G. The length of the single stranded DNA is the number of nucleotides comprising the single strand. The length of the double stranded DNA is counted in the number of base pairs. For more discussions of the relevant biological background refer [15, 16].

In the Adleman-Lipton model [1, 13], splints were used to construct the corresponding edges of a particular graph of paths, which represented all possible binary numbers.

Adleman and coworkers [14] proposed the sticker based model, which was an abstract of molecular computing based on DNA with a random access memory as well as a new form of encoding the information.

Adleman presented the new concept of computation in the molecular level in his work [1], and according to his ideas, we could build a molecular computer with the tools described as follows:

1. **Watson-Crick complements.** Two strands of DNAs will anneal to form the famous double helix, if the respective base meets its Watson-Crick complements.
2. **Ligases.** DNA molecules can be linked together through the process called ligation which mediated by enzymes called ligases. In a double stranded DNA, if one of the single strands contains a discontinuity then this may be repaired by DNA ligase.
3. **Amplification.** DNA molecules are amplified by the method polymerase chain reaction (PCR). PCR is a process that quickly amplifies the amount of DNA in a given solution.
4. **Gel electrophoresis.** A solution of DNA molecules is placed in one end of gel, and some electric current would be applied to the gel. This process is able to separate DNA strands by any specific length.
5. **DNA synthesis.** A desired strand of DNA can be synthesized using synthesizer in laboratory.

The synthesizer is supplied with the four nucleotide bases in solution, which are combined according to a sequence entered by the user. The instruments make millions of copies of the required oligonucleotides and place them in the solution.

The DNA operations in the Adleman-Lipton model are described below. These operations will be used for figuring out solutions of the independent dominating set problem.

**Operations on DNA molecules**

A test tube is a set of molecules of an alphabet {A, C, G, and T}. Given a tube, one can perform the following operations:

**Append.** Given a tube T and a short strand of DNA, Z, and the operation will append the short strand Z, onto the end of every strand in the tube T.

**Merge.** Given tubes T1 and T2, yield ∪(T1, T2) where ∪(T1, T2) = T1∪ T2. This operation is to pour two tubes into one, with no change of the individual strands.

**Copy (T, Ti).** In parallel, this operation produces a number of copies, Ti of the set T.

**Detect.** Given a tube T, say 'yes' if T includes at least one DNA molecule, and say 'no' if it contains none.

**Read.** Given a tube T, the operation is used to describe a single molecule, which is contained in the tube T.

Even if T contains many different molecules each encoding a different set of bases, the operation can give an explicit description of exactly one of them

**Extract.** Given a tube T and a short single strand of DNA, S, produces two tubes + (T, S) and – (T, S), where + (T, S) is all of the molecules of DNA in T which contain the strand S as a sub strand and – (T, S) is all of the molecules of DNA in T which do not contain the short strand S.

**Discard.** Given a tube T, the operation will discard the tube T.

**Sticker based solution space for the independent dominating set problem**

**Definition of the independent dominating set problem**

Let G be a graph with vertex set V (G) and the edge set E (G). A dominating set of a graph G = (V, E) is a subset D⊆V such that each vertex in V \ D is adjacent to at least one vertex in D. A

4. dominating set D is said to be an independent dominating set if no two vertices in D are adjacent in G. The independent dominating set problem is to find a minimum size independent dominating set in G and has been proved to be NP-Complete problem [3]. *D* is the independent domination number. For example, consider the graph G with 6 vertices and 9 edges.

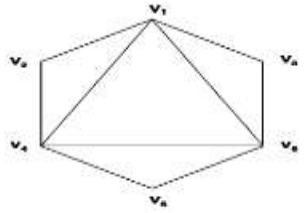

Fig 1.

**Fig 1.**

The minimum size dominating sets for the graph given in fig 1 are{ , } 1 6 *v v* , { , } 2 5 *v v* , { , } 3 4 *v v* , { , } 4 5 *v v* , { , } 1 5 *v v* , { , } 1 4 *v v* …. . The independent dominating sets for the given graph are { , }, 1 6 *v v* { , }, 2 5 *v v* { , }, 3 4 *v v* { , , } 2 3 6 *v v v* … The minimum size independent dominating sets are{ , } 2 5 *v v* , { , } 1 6 *v v* , { , } 3 4 *v v* . The independent domination number for the given graph is 2.

**Constructing the solution space of DNA sequences for the independent dominating set problem**

**Procedure:** The input is an undirected graph G = (V, E) where V is the set of vertices and E is the set of edges. *V* represents the number of vertices and *E* represents the number of edges in G. Let *V* = n and *E* = m. Initially, the test tube T1 contains n DNA sequences and each DNA sequence is a coding for each vertex *i v* , $1 \le i \le n$ and which is of the form 1 *i i v x* .

For the vertex *j v* that is adjacent to *i v* , Append ( , ) 1 1 *j T x*End ForAmplify ( ) 1 *T*

For *i* = 1 to n

T1=+(T1,vi)

T1=-(T1,vi)

Append ( , ) 1 *i T v*

Merge ( , ') 1 1 *T T*

End For

For generating the solution space of independent dominating set, we need n DNA sequence and each represents the vertex of the graph G and n DNA sequence to represent xj1 , j = 1 to n, the adjacency relation of the vertices of the given graph. To generate the solution space of the, the algorithm needs c+n append operations, one amplification operation, n merge

operation, where c is the sum of the degrees of the vertices of the graph G.

Now the test tube T1 contains n DNA sequences of the form v1 x11 x21 x31… xn1 v1,v2 x21x11 x31… xn-11 v2, …. vn xn 1 x11 x21… xn-11 vn. Next, generate the test tube T2, which contains the DNA sequence of the form $i\ j\ v\ v$ , where $v\ v\ E\ i\ j$ ( , )∉ . The initial set of DNA molecules encoding all the vertices and non edges of the given graph is synthesized using ABI 3948 nucleic acid synthesis and purification system. Consider the graph given in fig1. For each vertex $i\ v$ , i= 1 to 6 generate the DNA sequence 1 $i\ i\ v\ x$ . For representing the adjacency of the vertices of the graph, the DNA sequence xj 1, j = 1 to 6 are generated. The encoding scheme of adjacency relation of the vertices is given in the following table.

**The DNA algorithm for solving the independent dominating set problem**

Input ( , ) 1 2 $T\ T$ , where 1 $T$ includes all the encodings of the vertices with its adjacency relation and 2 $T$ includes the DNA sequence of the form $i\ j\ v\ v$ , where $v\ v\ E\ i\ j$ ( , )∉ .

**Step1.** Add multiple copies of DNA strands in the test tube 2 $T$ to the test tube 1 $T$ .

**Step2.** For each edge $v\ v\ E\ i\ j$ ( , )∈

( , ) 2 1 $i\ T = + T\ v$

$T = + T\ v$

Discard ( ) 3 $T$

End For

**Step3.** For $j$ = 1 to $n$

( , ) 1

1 1 $j\ T = + T\ x$

End For

**Step4.** If Detect ( ) 1 $T$ = yes then

4($a$) 1 $T$ = ( , ) 1 ∗ $T\ l$ , where $l$ is equal to the shortest length DNA strand in the Test tube 1 $T$

4($b$) Read ( ) 1 $T$

else

4($c$) independent dominating set is empty

End If

**Implementation of the algorithm**

This section describes the implementation of the algorithm.

• Start with two test tubes T1 and T2, T1 contains all the n DNA sequences which are the encodings of the n vertices with its adjacency relation and T2 contains DNA strands of the form $i\ j\ v\ v$ , where $v\ v\ E\ i\ j$ ( , )∉ .

• Add multiple copies of all the DNA encodings of T2 to the test tube T1 along with ligation enzyme. The ligation reaction which in turn produces the subsets of the vertex set V of the graph G. Since the test tube T2 contains the encodings of non edges of the graph G, each DNA strands in a test tube T1 is the DNA representation of subsets of the vertex set V of the given graph G with its adjacency relation. For each edge $v\ v\ E\ i\ j$ ( , )∈ , step2 checks whether the vertices vi and vj are in the same subset. If so, step2 removes those DNA strands from the test tube T1.

• The step 3 checks whether each vertex in V/S is adjacent to at least one vertex in S for each subset S of the vertex set V. The step 3 of the algorithm retains all the DNA strands in T1 which represent the independent dominating set of the problem using extraction operation. Thus, it needs n extraction operation.

• Detection operation is used to detect the content of the test tube T1. If it returns yes, because the independent dominating set problem is finding the minimum size independent dominating set, use step 4(a) to find the shortest length DNA strand from the test tube T1. This accomplished by gel electrophoresis technique. Use the read operation to find the solutions to the problem. Otherwise the independent dominating set is empty for the given graph G.

**Complexity analysis of the proposed DNA algorithm**

The following section describes the time complexity of the algorithm, the volume complexity of the solution space of the algorithm.

**Theroem1:** The independent dominating set problem for any undirected n-vertex graph G can be solved with $O(n + 2m)$ biological operations in the Adleman-Lipton model.

**Proof:** Algorithm can be applied for solving the independent dominating set problem for any undirected n vertex graph G. Algorithm given in this paper includes three main steps. Firstly, the algorithm given in step 1 generates the subsets of the vertex set V. Secondly; the step 2 extracts all the subsets of the vertex set V in which no two vertices are adjacent in the given graph G. Thus, the For loop given in step 2 uses 2m extraction operations. Step3 given in the algorithm checks whether each vertex in V/S is adjacent to at least one vertex in S for each subset of the vertex set V. It retains all the independent dominating set and uses n extraction operations.

Thirdly, the algorithm finds the minimal independent dominating set from the step 4(a) and uses read operation to find the solution of the problem. Hence from the statements mentioned above, it is inferred that the time complexity of the algorithm is $O(n + 2m)$ biological operations in the Adleman-Lipton model.

**Theorem2:** The independent dominating set problem for any undirected n vertices graph G can be solved with (2 ) 2 $O\ n + nC - m$ strands in the Adleman-Lipton model.

**Correctness of the Algorithm**

**Theorem:** From those steps given in the algorithm, the independent dominating set problem for any n-vertex graph can be solved.

**Proof:** The input of the algorithm is two test tubes T1 and T2. T1 includes all the n DNA strands which are the encoding for the n vertices with its adjacency and T2 includes all the DNA strands of the form $i\ j\ v\ v$ , where $v\ v\ E\ i\ j$ ( , )∉ . For any n vertex graph G, according to the first step of algorithm, it generates the subsets of the vertex set V of the graph G. The step 2 selects the subsets of the vertex set V in which no two vertices are adjacent in the given graph G. Now the test tube T1 contains the subsets of the graph G in which no two vertices are adjacent. Next step3 selects all the DNA strands which are the representation of subset S in which no two vertices are adjacent in G and each vertex not in the subset S is adjacent to at least one vertex in S. Because the independent dominating set problem is to find a minimum size independent dominating set, step 4(a) selects the shortest strand from the test tube T1 using gel electrophoresis and use the read operation to describe the sequence of molecules in the test tube T1. Otherwise independent dominating set is empty for the given graph G.

The graph given in fig 1 is used to show the power of the algorithm. Initially, the test tube T1 is filled with 6 library strands which represent the 6 vertices of the graph G with its adjacency relation given in Table1, section 3. The test tube T2 contains all the nC2-m DNA strands; each represents the non edge of the graph. That is the test tube T2 contains the DNA strands, $v\ v$ , 2 5 $v\ v$ , 2 6 $v\ v$ , 3 4 $v\ v$ 3 6 $v\ v$ . According to the first execution of step1, add multiple copies of all the DNA encodings in the test tube T2 to the test tube T1 with the ligation

enzyme. The ligation reaction produces the subsets{ , }, 1 6 *v v* { , }, 2 3 *v v* { , }, 2 5 *v v* { , }, 2 6 *v v* { , }, 3 4 *v v* { , }, 3 6 *v v*{ , , } 2 3 6 *v v v* , { , , } 2 3 4 *v v v* … of the given graph G with its adjacency relation in the test tube T1.

The step 2 selects all the subsets { , }, 1 6 *v v* { , }, 2 5 *v v* { , }, 3 4 *v v* { , , } 2 3 6 *v v v* … from the test tube T1, in which no two vertices are adjacent in G. For every DNA strand in the test tube T1, the step 3 checks whether other vertices not in the subset are adjacent to at least one vertex in the subset. Finally the test tube T1 contains all the DNA strands represent the independent dominating set. Because the test tube T1 is not empty, the detection operation for detecting the tube T1 in step 4 in the algorithm returns yes. Otherwise independent dominating set for the given graph is empty. The test tube T1contains all the independent dominating set, because independent dominating set problem is finding a minimum independent dominating set, the step 4(a) finds the shortest strand in the test tube T1. Therefore the step 4(b) reads the answer from the tube T1. Thus a minimum size independent dominating sets for the graph given in fig1 are{ , } 2 5 *v v* , { , } 1 6 *v v* , { , } 3 4 *v v* .

**Discussions and conclusions**

The importance of DNA computing technology in application has increased significantly in recent years. The proposed algorithm, demonstrates the independent dominating set problem can be solved using a DNA-based model with favorable computational efficiency. The approach described herein mostly transforms subsets of the vertex set of the graph into DNA strands, constructs the solution space, and then uses basic biological operations to generate a program that selects the qualified strands. Two major features of the proposed approach are as follows.

Firstly, it develops a computer program to generate a good DNA sequences for generating solution space of the independent dominating set problem with lower rate of errors. Secondly, the number of tubes needed for the algorithm will not increase with the complexity of the problem. The algorithm described in this paper is based on test tubes, and their computational power was greatly limited by the volume of DNA molecules which can be manipulated in the laboratory. As the complexity of the problem increases, longer DNA strands will have to be employed to encode the candidate solutions.

Since the completeness of the algorithm was verified, this framework suggests that algorithmic approaches that use biological operations to DNA strands may also be extended to solve more NP-complete problems.

**References**

1. L. Adleman, Molecular computation of solutions to combinatorial problems. Science, 266, 1994, 1021-1024.
2. K. H. Zimmermann, Efficient DNA sticker algorithms for NP-complete graph problems. Computer Physics Communications, 144, 2002, 297-309.
3. M. R. Garey, D. S. Johnson. Computers and Intractability: A Guide to the theory of NPcompleteness, Frema, San Francisco, 1979.
4. Q. Huiqin, L. Mingming, Z. Hong, Solve maximum clique problem by sticker model in DNA computing, Progress in natural science, 14, 2004, 1116-1121.
5. Wenbin Liu, Fengyne Zhang, Jin Xu, A DNA algorithm for the graph coloring problem, Journal of chemical and information science, 2002,42, 1176-1178.
6. K. L. Li, F. J. Yao, J. Xu, Improved molecular solutions for the knapsack problem on DNA based super computing, Chinese journal of computer research and development, 44, 2007, 1063-1070.
7. Sun-YuanH Sieh, Ming-yu Chen, A DNA based solution to the graph isomorphism problem using Adleman-Lipton model with stickers, Applied mathematics and computation, 197, 2008, 672-686.
8. Majid.D. A new solution for maximal clique problem based sticker model, Biosystems, 95, 2009, 145-149.
9. W. L. Chang, Solving the independent set problem in a DNA base super computer model, Physics review letter, 15, 2005, 469-79.
10. W. L. Chang, Minyi Guo, Michael Ho, Towards solution of the set splitting problem on gel based DNA computing, Future generation computer systems, 20, 2004, 875-885.
11. Minyi Guo, Michael Ho, W. L. Chang, Fast parallel solution to the dominating set problem on massively parallel biocomputing, parallel computing, 30, 2004, 1109-1125.
12. Feynman. R. P, in: Gilbert. D. H, Miniaturization. Reinhold publishing corporation, New York, 1961, 282-296.
13. Lipton. R. J, DNA solution of hard computational problems, Science, 268, 1995, 542- 545.
14. Roweis. S. et al. A sticker based model for DNA computation, Second annual workshop on DNA computing, DIMACS: Series in discrete and theoretical computer science, American mathematical society, Princeton university, 1999, 1-29.
15. Sinden . R. R, DNA structure and Function, Academic press, New York, 1994.
16. G. Paun, G. Rozenberg, A. Salomaa. DNA computing: New computing paradigms, springer verlag, New York, 1998.
17. EYES. Energy-Efficient Sensor Networks. (IST-2001-34734),http://www.eyes.eu.org.

**Table 1.**

| vertex | DNA Sequence |
|---|---|
| $V_1$ | $V_1 X_1^1 X_2^1 X_3^1 X_4^1 X_5^1 V_1$ |
| $V_2$ | $V_2 X_2^1 X_1^1 X_4^1 V_2$ |
| $V_3$ | $V_3 X_3^1 X_1^1 X_5^1 V_3$ |
| $V_4$ | $V_4 X_4^1 X_1^1 X_2^1 X_5^1 X_6^1 V_4$ |
| $V_5$ | $V_5 X_5^1 X_1^1 X_3^1 X_4^1 X_6^1 V_5$ |
| $V_6$ | $V_6 X_6^1 X_4^1 X_5^1 V_6$ |