Exploring Genetic Algorithm for Shortest Path Optimization in Data Networks

GJCST Classification (FOR)
C.2.1, G.2.2

Dr. Rakesh Kumar[1], Mahesh Kumar[2]

***Abstract* - Internet Service providers (ISPs) are the building blocks for Internet. Due to huge demand of Internet by various business communities and individuals, ISPs are trying to meet the increasing traffic demand with improved utilization of existing resources and application of new technologies. Routing of data packets can affect network resource utilization. A protocol is followed to send packets from source to destination along a path. In intra-domain Internet routing protocol Open Shortest Path First (OSPF) is the most commonly used protocol. As any user can come in and out from the logical topology of network, routing in dynamic network is a challenging one. I have implemented a Genetic algorithm to finds the set of optimal routes to send the traffic from source to destination.**

Keywords- Genetic Algorithm, Chromosome, Crossover, Mutation, Routing

I. Introduction

Data network routing is a process of transferring packets from source node to destination node with minimum cost (delay-transmission, processing and queuing delay, bandwidth, load, jitter, reliability etc.). Routing is complex in large networks because of the many potential intermediate destinations a packet might traverse before reaching its destination. Hence routing algorithm has to acquire, organize and distribute information about network states. It should generate feasible routes between nodes and send traffic along the selected path and also achieve high performance. Routing in conjunction with congestion control and admission control defines the performance of the network. The weights of links in network are assigned by the network operator. The lower the weight, the greater the chance that traffic will get routed on that link [BLU00]. When one sends or receives data over the Internet, the information is divided into small chunks called packets or datagram's. A header, containing the necessary transmission information, such as the destination Internet Protocol (IP) address, is attached to each packet. The data packets are sent along links between routers on Internet. When a data packet reaches a router, the incoming datagram's are stored in a queue to await processing. The router reads the datagram header, takes the IP destination address and determines the best way to forward this packet for it to reach its final destination. Routing algorithm should havegeneric objective ofrouting strategy to be both dynamically reconfigurable and be based on locally available information. It should also satisfy user quality of service objectives. Some of the methods proposed in achieving these objectives are social evolutionary algorithms, insect metaphors and cognitive packet network. Insect metaphors and cognitive packet network methods use the probabilistic routing table and allow the packets themselves to investigate, report network topology status and performance. Genetic algorithm is one in which the population associated with each node co-evolve to solve the problem.

About[1] : Associate Professor, Department of Computer Science and Applications, Kurukshetra University Kurukshetra, India (rsagwal@gmail.com, 9896336145)

About[2] : Assistant professor, Department of Computer Science and Engineering, DAV college of engineering & technology, Kanina, Mohindergarh, India (myadav.kuk@gmail.com, 9813431025)

II. Network Routing

Routing is a process of finding paths between nodes in network. Broadly there are mainly two types of routing policies - static and dynamic. In static routing, the routes between the nodes are precomputed based on certain factors for example bandwidth, buffer space etc. and are stored in routing table. All packets between any two nodes follow the same path. When network topology changes the path between two nodes may also change and static routing fails. Hence in dynamic routing policy, the routes are not stored but are generated when required. The new routes are generated based on the factors like traffic, link utilization, bandwidth, jitter, delay etc which is aimed at having maximum performance. For message transmission routing policy may be centralized or distributed. In the case of centralized routing, only centralized node generates routes between any pair of nodes. Centralized routing is not adequate in IP networks as it is required to collect whole network state before route computation, which is very complex task. In distributed routing, each node generates routes independently between pair of nodes as and when required. Other classification of routing policy is optimal routing (global routing) and shortest path routing (local routing) [CAUOO]. Some of the shortest path algorithms are distance vector algorithm and link state algorithm. Each node in the network is of the type store and forward. The link performance may be measured in terms of bandwidth or link delay. The topology of the network may change due to growth in number of nodes, or failure of node. This change in topology should be reflected in the routing table, which in turn helps the routing protocol to generate optimal route for the current state of network. Some of the protocols are Routing Information Protocol (RIP), Interior gateway routing protocol (IGRP), Open source shortest path first (OSPF) and Border gateway protocol (BGP). OSPF is a link state routing protocol used in IP networks which uses shortest path first algorithm to compute low cost route to destination. Enhanced Interior Gateway Routing Protocol (EIGRP) is an enhanced distance vector routing protocol with optimization to minimize the effect of change in topology and efficient use of bandwidth and processing power at the router. EIGRP uses unequal cost load balancing. Routing process uses routing table at each node to store all the nodes which are at one hop distance from it [CAU00]. It also stores the other nodes (hop count more than one) along with the number of hops to reach that node, followed by the neighbor node through which it can be reached. Router decides which neighbor to choose from routing table to reach specific destination.

III. Shortest Path Problem

The shortest path problem is defined as that of finding a minimum-length (cost) path between a given pair of nodes. The Dijkstra algorithm is considered as the most efficient method for shortest path computation in IP networks. But when the network is very big, then it becomes inefficient since a lot of computations need to be repeated [BIL08]. Also it cannot be implemented in the permitted time [YIN06].

IV. Genetic Algorithm

Genetic Algorithm (GA) is a special kind of stochastic search algorithm that depicts the biological evolution as the problem solving technique. GA works on the search space called population [GOL89]. Each element in the population is called as chromosome. GA begins with randomly selecting set of feasible solution from population. Each chromosome is a solution by itself. Each chromosome is evaluated for fitness and this fitness defines the quality of solution. GA uses adaptive heuristic search technique which finds the set of best solution from the population. New off-springs are generated /evolved from the chromosomes using operators like selection, crossover and mutation. Most fit chromosomes are moved to next generation. The weaker candidates get less chance for moving to next generation. This is because GA is based on the principle of Darwin theory of evolution, which states that the "survival is the best". This process repeats until the chromosomes have best fit solution to the given problem [LAU91]. The summary is that the average fitness of the population increases at each iteration, so by repeating the process for many iterations, better results are discovered. GA has been widely studied and experimented on many fields of engineering. GA provides alternative methods for solving problems which are difficult to solve using traditional methods. GA can be applied for nonlinear programming like traveling salesman problem, minimum spanning tree, scheduling problem and many others.

1) Features of Genetic algorithm

• The most important feature of genetic algorithms is that they are parallel in nature. They explore solution space in multiple directions at once. GA is well suited for solving problems where the solution space is huge and time taken to search exhaustively is very high.

• GA performs well in problems with complex fitness. If the function is discontinuous, noisy, changes over time or has many local optima, then GA gives better results [VIJ08].

• GA has ability to solve problems with no previous knowledge (blind). The performance of GA is based on efficient representation, evaluation of fitness function and other parameters like size of population, rate of crossover and mutation and the strength of selection.

2) Problem definition

The network under consideration is represented as G = (V, E), a connected graph with N nodes. The metric of optimization is cost of path between the nodes. The total cost is the sum of cost of individual hops. The goal is to find the path with minimum total cost between source node V_{src} and destination V_{dest}, where V_{src} and V_{dest} belong to V. This paper presents the efficient on-demand, source initiated routing algorithm using genetic algorithm. Finally data is sent along the generated path.

A. Initialization of routing table

A module is used to generate all possible paths from a given node to all other nodes in the network. Initially, „n' random paths are considered (chromosome). This 'n' defines the population size. These chromosomes act as population of first generation.

B. Optimal paths generation

This module deals with finding the optimal path using genetic algorithm. The input to this module is the set of paths generated. Each path is called as chromosome. As the source node receives „m' (say 10- population size) chromosomes-

(a) Calculate the fitness of each of the chromosome.

The fitness of the chromosome is evaluated as: Fitness = no of hops in path * 10 – total cost of path Number of hops defines the number of intermediate nodes visited along the path from source to destination and total cost is the sum of cost of individual links in the path.

(b) Select best two chromosomes as parents (using some selection method-Roulette wheel)

(c) Perform crossover with probability 0.7

(d) Perform mutation with probability 0.01

(e) Place children in the population and eliminate the worst chromosome having very poor fitness value.

(f) If termination condition is not attained then repeat the steps i. to vi.

Else (Convergence criteria is reached) { • Store the paths for duration t seconds • send data to the destination along the path }

(g) Refresh the path after duration of t seconds to know the current status of dynamic network.

C. Selection

It is a feature of GA for selecting parents for next generation. Current work is based on roulette wheel selection.Roulette wheel selection – In roulette wheel selection, the individual is selected based on the relative fitness with its competitors. This is similar to dividing the wheel into a number of slices. Fittest chromosomes get larger slice. Some of the other selection methods are rank selection, elitist selection, scaling selection, tournament selection, etc.

D. Crossover

Crossover operator combines sub parts of two parent chromosomes and produces offspring that contains some part of both the parent genetic material. Crossover is mainly of two types namely single point crossover and multipoint crossover. In single point crossover, there is one cross over site and in multipoint crossover there is more than one crossover site. Single point crossover method is simple; it has some problems like formation of cycles when used for routing. Hence it is required to use some of the advanced multipoint crossover techniques to eliminate cycle. Some of the advanced multipoint crossover techniques are Partially Mapped Crossover (PMX), Cycle crossover (CX) and Order crossover (OX) [GOL98, SIV08]. This paper deals with PMX crossover. In Partially Matched Crossover [SIV08], two strings are aligned, and two crossover points are selected uniformly at random along the length of the strings. The two crossover points give a matching selection, which is used to affect across through position-by-position exchange operations.

Consider two strings:

Parent A 4 8 7 | 3 6 5 | 1 10 9 2
Parent B 3 1 4 | 2 7 9 | 10 8 6 5

Two crossover points were selected at random, and PMX proceeds by position wise exchanges. In-between the crossover points the genes get exchanged i.e., the 3 and the 2, the 6 and the 7, the 5 and the 9 exchange places. This is by mapping parent B to parent A. Now mapping parent A to parent B, the 7 and the 6, the 9 and the 5, the 2 and the 3 exchange places. Thus after PMX, the offspring produced as follows:

Child A 4 8 6 | 2 7 9 | 1 10 5 3
Child B 2 1 4 | 3 6 5 | 10 8 7 9

Each offspring contains ordering information partially determined by each of its parents. PMX can be applied to problems with permutation representation. Generated offspring should be validated. Validation is done by checking the offspring with all possible routes. If offspring belongs to all possible routes then its fitness is computed and sent to next operation. If the offspring does not belong to all possible route set, then it is dropped as route does not involve valid connections of nodes in network.

E. Mutation

Crossover operation may produce degenerate population. In order to undo this, mutation operation is performed. Mutation operation can be bit flipping, interchanging, inversion, insertion, reciprocal exchange or others [ALU06]. The paper uses insertion method. In case of insertion a node is inserted at random position in the string. This is because a node along the optimal path may be eliminated through crossover. Using insertion, it can be brought back. Once mutation is completed, the offspring generated by mutation have to be validated with the same process used in crossover.

F. Termination Criteria

It allows the convergence of algorithm. Maximum generations, No change in population fitness and stall generation are considered as algorithm stopping condition. We have taken the maximum number (say 1000) of generations as it will allow algorithm to check, upto what number of generations there is improvement in chromosome fitness. A second stopping criterion is until some chromosome reaches a specified fitness level. As the optimal solution is generated using GA, data is transmitted along that path. There may be change in topology of network as some nodes may join the network or some nodes may leave the network or some nodes may fail. Under these circumstances the optimal path may no more be the shortest. Hence the network has to be refreshed at every t seconds and new routes may be generated.

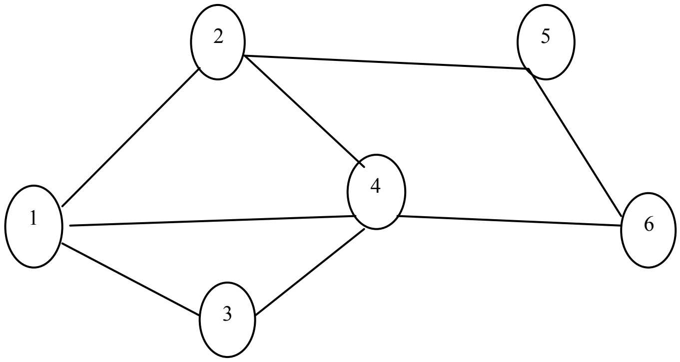

Figure 1 : Sample Network Topology

The Cost on links is given in table 1:

	1	2	3	4	5	6
1	999	**5**	**3**	**7**	999	999
2	**5**	999	999	**3**	**5**	999
3	**3**	999	999	**3**	999	999
4	**7**	**3**	**3**	999	999	999
5	999	**5**	999	999	999	**3**
6	999	999	999	**2**	**3**	999

Table 1: Cost on linksThe values 999 represents that there is no direct link between these nodes. 999 is a big value as compared to other costs. During implementation only small values are considered for path computation.

V. Results

Current work is based on network consisting 6 nodes. Initially 15 random chromosomes are generated, out of which best ten are considered for 1st generation. At each generation the chromosomes are validated and best fit chromosomes are sent to next generation. It is found that fitness value improves at each generation for chromosomes.

Generate 15 random chromosomes

Chromosome	Delay	Fitness
1 4 2 5 6 3	18	32
1 2 5 6 4 3	18	32
1 4 6 5 3 2	12	18
1 2 4 3 5 6	11	19
1 2 4 3 6 5	11	19
1 2 5 4 3 6	10	10
1 4 2 3 5 6	10	10
1 4 6 3 5 2	9	11
1 2 4 5 3 6	8	12
1 2 3 4 5 6	5	5
1 2 3 6 4 5	5	5
1 2 6 4 5 3	5	5
1 2 6 4 3 5	5	5
1 3 5 6 4 2	3	7
1 6 3 4 5 2	0	0

Generation 1

Chromosome	Delay	Fitness	Nodes visited
1 4 2 5 6 3	18	32	5
1 2 5 6 4 3	18	32	5
1 4 6 5 3 2	12	18	4
1 2 4 3 5 6	11	19	4
1 2 4 3 6 5	11	19	4
1 2 5 4 3 6	10	10	3
1 4 2 3 5 6	10	10	3
1 4 6 3 5 2	9	11	3
1 2 4 5 3 6	8	12	3
1 2 3 4 5 6	5	5	2

Generation 2

Chromosome	Delay	Fitness	Nodes visited
1 3 4 2 5 6	17	33	5
1 3 4 6 5 2	16	34	5
1 4 6 5 3 2	12	18	4
1 2 4 3 5 6	11	19	4
1 2 4 3 6 5	11	19	4
1 2 5 4 3 6	10	10	3
1 4 2 3 5 6	10	10	3
1 4 6 3 5 2	9	11	3
1 2 4 5 3 6	8	12	3
1 2 3 4 5 6	5	5	2

We have taken population size of 10 in first generation. By selecting the chromosomes based on roulette selection and application of GA operators generations are performed. After the path to all nodes from source node 1 is computed, the set of paths to a specific node will be displayed. Let the destination node is node 6. Following is the set of paths from node 1 to node 6. The optimal path returned is 1, 3, 4, and 6 with delay factor of 8.

Source	Destination	Delay	Route
1	6	13	1 2 5 6
1	6	10	1 2 4 6
1	6	9	1 4 6
1	6	18	1 4 2 5 6
1	6	8	1 3 4 6

Table2: Routes to the destination 6

VI. Conclusion

Genetic algorithm is used for routing in packet switched data networks in current work. They explore solution space in multiple directions at once. GA is well suited for solving problems where the solution space is huge and time taken to search exhaustively is very high. As the size of network increases, the possible solutions for transferring data between two nodes increase. Adding of few new nodes in the network increases the size of search space exponentially. So, GA is well suited for routing problem as it explores solution space in multiple directions at once and less chances to attain local optimum.GA has ability to solve problems with no previous knowledge. The performance of GA is based on efficient representation, evaluation of fitness function, population size, crossover rate, mutation probability and the selection method. The proposed algorithm works on initial population created by some other module, access fitness, generate new population using genetic operators and converges after meeting to specified termination condition. Current work can be improved by using some intelligent approach for populating routing table and using better crossover, mutation probabilities and enhancing it to support for load balancing.

VII. Refrences

1) (ALU06) Aluizio F. R. Araújo et. al. (2006): Multicast Routing Using Genetic Algorithm Seen as a Permutation Problem, Proceedings of the 20th International Conference on Advanced Information Networking and Applications (AINA'06)
2) [BIL08) Bilal Gonen (2008): Genetic Algorithm Finding the Shortest Path in Networks, 2008
3) [BLU00) Black U. (2000): IP Routing Protocols, RIP, OSPF, BGP, PNNI & Cisco routing protocols, Prentice Hall
4) (CAU00) Cauvery, N. K. & Viswanatha, K. V. (2000): Routing in Dynamic Network using Ants and Genetic Algorithm, International Journal of Computer Science and Network Security, VOL.9 No.3, March 2000
5) (GOL89) Goldberg E. David (1989): Genetic Algorithms in search, optimization and machine learning, Pearson Education
6) (LAU91) Laurence Davis (1991): Hand book of Genetic Algorithms, 1991
7) [SIV08) Sivanandam S.N. & Deepa S.N. (2008): Introduction to Genetic Algorithms, 2008

8) (VIJ08) Vijayalakshmi K & Radhakrishnan S (2008): Dynamic Routing to Multiple Destinations in IP Networking using Hybrid Genetic Algorithm (DRHGA), International Journal of Information Technology, Vol 4, No 1, PP 45-52, 2008
9) (YIN06) Yinzhen Li, Ruichun He, & Yaohuang Guo. (2006): Faster Genetic Algorithm for Network Paths, The Sixth International Symposiumon Operations Research and Its Applications (ISORA'06), pp.380–389