A Genetic Algorithm for the Overlay Multicast Routing Problem

PAN Yun[1] YU Zhenwei[1] WANG Licheng[2]
(1 Dept. Computer, China University of Mining and Technology-Beijing, Beijing, 100083)
(2 Knowledge Engineering Group of Dept. Computer of Tsinghua University, Beijing 100084)

Abstract In this paper, we analyze several models of overlay multicast routing problem, and bring forward a new model based on multi-objective programming, discussing the solution of the model simultaneously. Then we employ the Prüfer sequence as chromosome code and then propose a genetic algorithm to solve the model. Finally, we analyze the complexity of the algorithm.
Key words overlay multicast, genetic algorithms, multicast routing

1 Introduction

Most of Internet applications are based on one-to-one unicast transmission, however, with the develop of Internet, many new applications, such as videos conference, remote education, CSCW and so on, need group communication service. In contrast to many one-to-one unicast to support group communication, multicast[1] is an efficient transmission mechanism. Multicast services allow one host to send information to a large number of receivers according to a multicast distribution tree, without network interface-bandwidth-constrained. This makes applications more scalable and leads to more efficient use of network resources. For realizing multicast, network needs to provide new routing policy and new transmission mechanism.

Despite the conceptual simplicity of IP multicast and its obvious benefits, it has not been widely deployed. There remains many unresolved issues[2] in the IP multicast model that hinder the development and deployment of IP multicast and multicast applications. The most prominent issues are the lack of a multicast address allocation scheme, the lack of access control and the lack of an inter-domain multicast routing protocol. Moreover, the worst drawback is that IP multicast mechanism ignore ISP's profit and then impede the development of multicasting application in lager scale.

As an alternative to IP multicast, the concept of overlay multicast[3,4,5,6] is brought forward recently. Overlay multicast uses the Internet as a low level of infrastructure to provide multicast service to end hosts. The strategy of overlay multicast avoid most of the basic deployment issues associated with IP multicast, such as end-to-end reliability, flow and congestion control schemes, unique address for each multicasting group, etc. Since the overlay multicast is built on the application level, it can provide significant flexibility to satisfy different application demands.

Typically, overlay multicast networks are built on top of a general Internet unicast infrastructure rather than point-to-point links, the problem of managing their resource usage is somewhat different than in networks that do have their own links, leading to differences in how they are best configured and operated[7]. First, conventional IP multicast routing algorithm uses short path tree, but overlay multicast networks adopt different routing algorithms to different applications; Second, IP multicasting route is constructed on top of real physics link, but overlay multicast routing is constructed on top of visual network. In overlay networks, a multicasting tree is a spanning tree on a complete graph because every pair of nodes is adjacent logically.

Current projects in overlay multicast propose several design approaches with different design objectives[9]. There are many typical schemes such as Narada[3]、Scattercast[4]、Overcast[5] and ALMI[6]. Narada and Scattercast intend to minimizing delay for each member, the objective of Overcast is to maximize available bandwidth for each member, ALMI strives to minimize the system cost, that is, improving the utility of networks.

In the second section of this paper, we take ALMI[6] as reference and put forward a rational multicast routing model based on multi-objectives programming; Then, we discuss the methods to solve the new model in section 3 and design a novel genetic algorithm for the model in section 4. Finally, The validity of our method is suggested by a large number of numerical experiments in section 5.

2 Overlay multicast Routing Problem

Overlay multicast routing algorithms have two main performance objectives[7,11]: First, routing algorithms must use network resources efficiently to carry as much traffic as possible. Second, routing algorithms must minimize end-to-end delay. However, these two objectives are two aspect of a tradeoff because a smaller end-to-end delay multicasting tree makes center nodes traffic concentration then these nodes become system bottleneck. On the other hand, increasing overall network usage results in distribute load on several different nodes, then this makes path longer and delay more. So, how to optimize the two parameters is a crucial problem.

An overlay multicast network can be modeled as a complete graph, because there exists a unicast path between every pair of nodes. For each multicast session all nodes share a multicast tree which can provide multicast service for all uses. Therefore, finding an overlay multicast tree needs to consider every node's interface-bandwidth-constrained. We employ the degree-constrained of node to denote the amount of interface-bandwidth-constrained in use to support a multicast session.

2.1 Delay Optimization

On condition of satisfying interface-bandwidth-constrained, minimizing end-to-end delay comes down to find a minimum diameter spanning tree subject to degree-constraints.

MDDL（Minimum diameter, degree-limited spanning tree）problem[7,11]：Given an undirected complete graph $G=(V,E)$,where *V* is a set of application level nodes and *E* is a set of edges. The cost of edge *e* is denoted by *c(e)*,which is a positive real number. The degree-constraint of each node *i* is denoted by $d_{\max}(v)\in N$. Then, the objective is to find a multicasting tree such that its diameter is minimum among all possible choices of spanning trees satisfying the degree-constraints, as follows:

$$\begin{cases} \min\limits_{T\subseteq G} \quad dia(T) \\ s.t. \quad d_T(v)\le d_{\max}(v), \forall v\in V_T \end{cases} \tag{1}$$

Where *T* is a spanning tree *G* , *dia(T)* is defined as the longest of the shortest paths in *T* among all the pairs of nodes in *V* and V_T denotes node set of *T*.

2.2 Load Balancing

MDDL can set up a minimum delay multicast tree. However, lack of load balancing, **MDDL**

causes system utilize ratio decrease and session refuse ratio raise. So [7] also puts forward load balancing way.

LDRB (Limited diameter, residual-balanced spanning tree problem) problem[7,11] :Given an undirected complete graph $G=(V,E)$,where *V* is a set of application level nodes and *E* is a set of edges. The cost of edge *e* is denoted by *c(e)*, which is a positive real number. The degree-constraint of each node *i* is denoted by $d_{\max}(v) \in N$. $B \in Z^{+}$ is diameter bound. Then, the objective is to find a multicasting tree such that its minimum residual degree $d_{\max}(v) - d_T(v)$ reaches maximum among all possible choices of spanning trees satisfying the degree-constraints and the diameter bound *B*, as follows:

$$\begin{cases} \max\limits_{T \subseteq G} \min(d_{\max}(v) - d_T(v)) \\ s.t. \quad dia(T) < B \\ d_T(v) \le d_{\max}(v), \forall v \in V_T \end{cases} \tag{2}$$

Where *T* is a spanning tree *G* , *dia(T)* is defined as the longest of the shortest paths in *T* among all the pairs of nodes in *V* and V_T denotes node set of *T*.

2.3 Multi-constrained routing scheme

Above two models optimize one objective on condition of fixing other objective as constraint. Moreover, **LDRB**'s load balancing way is very complicated, belongs to "maximum - minimum" optimal problems with high computational complexity. Therefore, we improve **LDRB**'s load balancing mechanism. It is clear that the regular graph is corresponding to the best of load balancing on condition of no degree-constrained. So, in overlay multicast routing problem every node has extent degree-constrained, the multicast tree with the smallest variance of δ_T is corresponding to the best of load balancing tree. Where $\delta_T(v) = d_{\max}(v) - d_T(v)$ is the residual vector of degree-constraints for a give multicast tree *T*. Now we consider optimal two parameter-delay and load balancing, application-level multicast routing problem is described as follows:

MDRBDL (Minimum diameter, residual-balanced,degree-limited spanning tree) problem : Given an undirected complete graph $G=(V,E)$,where *V* is a set of application level nodes and *E* is a set of edges. The cost of edge *e* is denoted by *c(e)*,which is a positive real number. The degree-constraint of each node *i* is denoted by $d_{\max}(v) \in N$. Our objective is to find a multicasting tree such that its diameter is minimum and its load balancing variance is minimum among all possible choices of spanning trees satisfying the degree-constraints, as follows:

$$\begin{cases} \min\limits_{T \subseteq G} \ dia(T) \\ \min\limits_{T \subseteq G} \ \delta(T) \\ s.t. \quad d_T(v) \le d_{\max}(v), \forall d \in V_T \end{cases} \tag{3}$$

Where *T* is a spanning tree *G* , *dia(T)* is defined as the longest of the shortest paths in *T* among all the pairs of nodes in *V* , V_T denotes node set of *T* and δ_T is the degree of *T* load-balancing.

2.4 The Complexity of **MDRBDL**

Lemma 1 The degree-constrained spanning tree problem is NP-Completed.

Instance：Given a graph $G = (V, E)$ and $d \in D^{|V|}, D = \{1,2,...,|V|-1\}$。

Problem：Is there a spanning tree of G such that the degree of its j^{th} node no more than d_j(j=1,2,⋯,|V|)?

Proof：Supposing d_j=2(j=1,2,…,|V|)，we can confined this problem to be a Hamilton path problem[8]. This is the end of the proof.

Lemma 2 The degree-constrained minimal spanning tree problem is NP-hard.

Proof: Because its decision problem is NP-Completed.

Theorem 1 MDRBDL problem is NP-hard。

Proof: Because its sub problem NP-hard.

Genetic algorithms (GAs) are random, adaptive search methods inspired by Darwin's theory about evolution. Solution to a problem solved by genetic algorithms is evolved. Since GA had been proposed by Holland in 1975, it has excellent performances in many fields, especially in solving some NP-hard optimization problems. We apply GA to solve the overlay multicast routing problems.

3 The Method to Solve MDRBDL Model

MDRBDL is a multi-objective programming problem. One of the methods to solve multi-objectives programming model is to convert it into a single objective programming model by introducing partial factor for each objective . Then (3) can be converted into (4)

$$\begin{cases} \min\limits_{T\subseteq G} \ g = \lambda_1 f_1 + \lambda_2 f_2 \\ s.t.\ d_T(v) \le d_{\max}(v), \forall v \in V_T \end{cases} \tag{4}$$

Where f_i and $\lambda_i \ge 0 (i = 1,2)$ respectively denote each objective function in (3) and the partial factor of f_i, and $\sum_i \lambda_i = 1 (i = 1,2)$ holds. In our experiment, we set partial factors λ_1=0.5, λ_2=0.5.

Furthermore, we apply penalty function method to transforming (4) into an unconstrained optimization model (5)

$$\min_{T\subseteq G} \ g = \lambda_1 f_1 + \lambda_2 f_2 + \left\|\vec{p}\right\| \tag{5}$$

Where $\|\cdot\|$ is a form which can be chosen according concrete applications. We use 2-form here. And the penalty vector $\vec{p} = (p_1, p_2, ..., p_n), n = |V|$ is defined as

$$p_j = \begin{cases} (1+\alpha)^{k + d_T(v_j) - d_{\max}(v_j)} \log_{\max G} k, d_T(v_j) > d_{\max}(v_j), v_j \in V_T \\ 0, others \end{cases} \tag{6}$$

Where k is the current evolution generation while max$G$ the maximum generation. And $\alpha \in (0,1]$ need to be chosen carefully. Because the method will be lack of penalty if α are too small; On the contrary, it will be over penalty if they are too lager. Other adaptive methods can also be used to choose α . We set $\alpha = 0.05$ according to a great number of random numerical experiments. Obviously, the penalty functions are dynamic. They will gradually strengthen penalty to infeasible solutions with the increasing of *k*.

4 Genetic Algorithm for Solving MDRBDL

4.1 The Prüfer sequences

Because the network graph of overlay multicast is a complete graph, multicasting tree is a spanning tree on the complete graph; at the same time the Prüfer coding is aim at complete graph, so we employ the Prüfer coding as the Prüfer coding sequences. There are n^{n-2} distinct sequences (called Prüfer sequences, or Prüfer codes) of length *n-2* with entries being from natural number 1 to *n*; Prüfer [1918 year] established a bijection between the set of Prüfer sequences and the set of spanning trees of complete graph K_n [10]. The Prüfer sequences provide one of the most concise encoding methods for designing genetic algorithms to solve the optimization problems that are correlated with spanning trees. The decoding algorithm is labeled *seqToTree* as follows[10]：

Algorithm seqToTree

Input: P, the Prüfer sequence P;

Output: T, the spanning tree of K_n;

[Start]

1） Set n: = length (P) + 2;

2） Count the appearing times A_i of the vertex i in P;

3） Set T is the graph with n isolated vertices which are labeled as 1,2…n respectively;

4） Set Q is the sorted sequence of T's vertices which are not in P;

5） While not empty (P) do

a） Remove the first element v from P and the first element u from Q,

b） Set A_v: =A_v –1;

c） Add the edge <v, u> to T;

d） if A_v =0 then insert v into Q(using binary-search method);

6） Remove the remained vertices v and u from Q, then Add the edge <v, u> to T;

7） Output T.

[End]

A Prüfer sequence denotes a free spanning tree of the complete graph K_n. It is different from a multicast tree. The latter needs a labeled node as root. So we expand the length of a Prüfer sequence to *n*-1 and let the last entry denote root.

4.2 The Design of the Genetic Algorithm

A genetic algorithm consists of several crucial aspects, such as chromosome encoding and decoding, population initializing, individual's fitness value computing and scaling, and genetic operating (selecting, crossing and mutating), and so forth.

Based on the discussion above, we designed a novel genetic algorithm to solve MDRBDL. Our method can be described as follows:

1) *Population initializing* Let *P*, a random matrix of $p\times(n-1)$, denote the population of individuals, where *p* is the population size while *n* the number of nodes of the given network *G*.
2) *Chromosome encoding* Let an extended Prüfer sequence, i.e. a row of the matrix *P*

represent a chromosome.

3) *Chromosome decoding* First, we use the algorithm *seqToTree* to decode a chromosome code, i.e. an extended Prüfer sequence to a spanning tree of completed graph K_n.
4) *Individual fitness value computing and scaling* We employ the formula (5) to estimate the fitness value, denoted by g, of the multicasting tree which is obtained in step 3) and then scaling g to f by the formula $f = \frac{1}{1+g}$.
5) *Genetic operating* We adopt an improved roulette wheel selection operator, a random multi-points crossover operator, and three mutation operators: a single gene mutating operator, a gene fragment reversing operator and a gene fragment shifting operator in the new algorithm.
6) *Elitism strategy* In each population, the best individual, called elitist, is kept from crossing and mutating.

4.3 Algorithm Analysis

According to the algorithm *seqToTree*, the complexities of decoding a Prüfer sequence to a spanning tree is $O(nlogn)$. As for the MDRBDL model, the complexity of computing and scaling the fitness is $O(n)$. All the complexities of crossing operator and mutating operators do not exceed $O(n)$.In conclusion, the total complexity of the new genetic algorithm, given the population size p and the maximum generation G, is $O(p\times \max G\times nlogn)$.

5 Simulation and Analysis

In order to validate our algorithm, we construct stochastically complete graph K_n and set the link cost to a random number. The degree constraint of each node is assigned to be equal to or smaller than its real degree value in the graph. We analyze the two aspect of the performance of our algorithm: 1)The constringency; 2) The quality of multicasting tree.

In the paper, we adopt adaptive crossing probability $p_c=1\text{-}md$ and mutating probability $p_m=md/20$. The gene fragment reversing probability p_r and gene fragment shifting probability p_s are set as 0.005. The algorithm in Fig 1 determines the maturation degree of current generation.

% Maturation degree of P
xstd=std(P);
vx=abs(mean(P));
svx=mean(xstd./(0.0001)+vx));
*md=1-(1+tanh(6*svx-3))/2;*

Fig. 1 maturation degree

First, we give a typical example as Fig.2, which shows a network graph of 6 nodes numbered 1 through 6. The degree-constraint of each node is given in each bracket and the communication cost of each pair nodes is also labeled on each link. Fig.3 gives the degree-constrained minimum diameter spanning tree for the example of Fig.2. The minimum diameter in the spanning tree is 8, and the extent of load-balancing is 1.067.

Second, Fig.4 shows the relation between evolution and cost(diameter and load-balancing) in a 10-node network graph. we can find that the algorithm have better constringency.

Finally, we set maximum generation $\max G=300$, population size $pop=40$, and run our algorithm 100 times, comparing our algorithm performance. In our experiment, we compare the diameter and load-balancing of multicasting tree with optional value according to the nodes change form 10 to 30. Fig.5 shows the comparison of the results obtained by the proposed algorithm and optional value, the diameter of multicasting tree is bigger than optional value, the extent of load balancing is bigger than optional value.

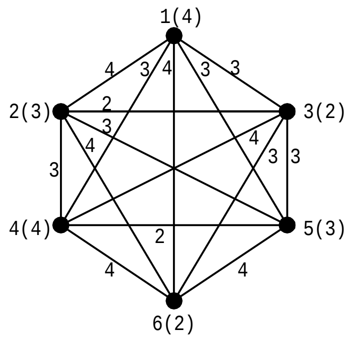

Fig.2 A example network with 6 nodes

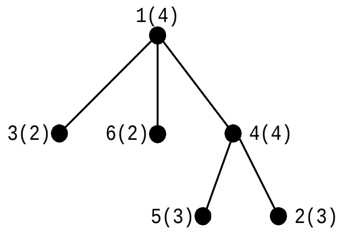

Fig.3 Result of applying GA to the example in Fig.2

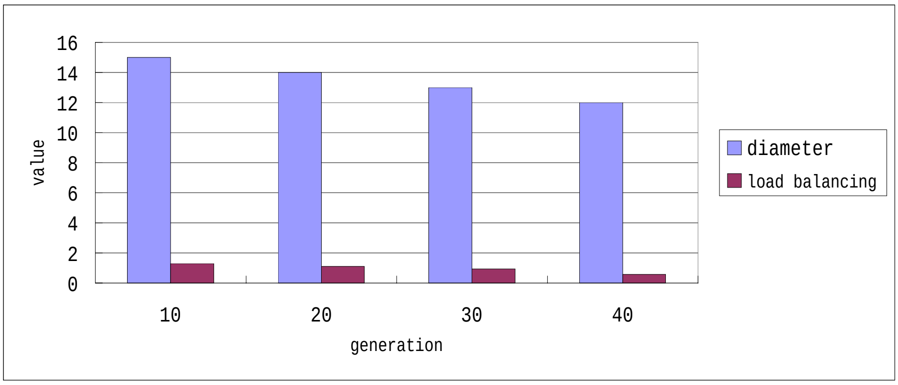

Fig.4 the relation between evolution generation and cost in a 10-node network graph

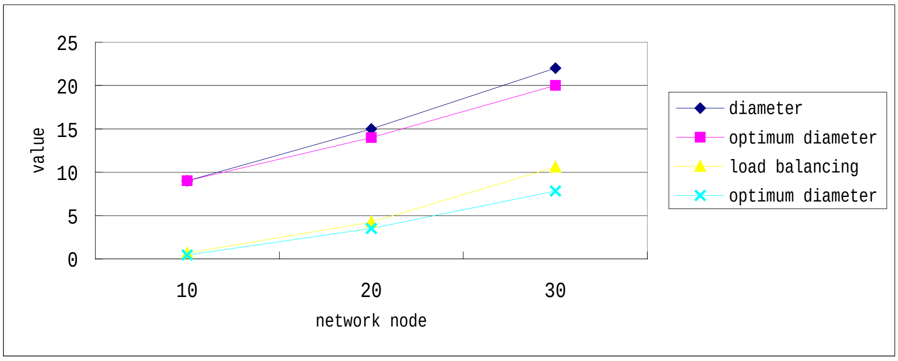

Fig.5 compare the diameter and load-balancing of multicasting tree with optimum value

6 Conclusion

It is a tradeoff between two objectives in multicast routing on overlay networks. The one is to balance the load of each node in the networks as much as possible. The other is to decrease transmission delay in best effort. MDDL and LDRB are two famous methods for them respectively. But all of these method are based on single objective programming model which optimize one objective and take another as constraint. In order to improve the performance of multicast routing in further degree, we construct a new model for the problem based on multi-objectives programming. And then, we design a novel genetic algorithm using an improved Prüfer coding method to solve the new model. The complexities of coding and decoding of our algorithm do not exceed $O(nlogn)$. A great number of numerical experiments suggest that our algorithm is feasible and efficient.

References

[1] DEERING, S. E., Multicast Routing in a Datagram Internetwork. PhD thesis, Stanford University, Dec. 1991.

[2] C. Diot, B. Levine, J. Lyles, H. Kassem, and D. Balensiefen, "Deployment issues for the IP multicast service and architecture," IEEE Network, 2000..

[3] Y. Chu, S. Rao, and H. Zhang, “A case for end system multicast”, in ACM Sigmetrics, 2000.

[4]Y. Chawathe, “Scattercast: An Architecture for Internet Broadcast Distribution as an Infrastructure Service”. PhD thesis, Department of EECS, UC Berkeley, Dec. 2000.

[5] J. Jannotti, D. K. Gifford, K. L. Johnson, M. F. Kaashoek, and J. W. O. Jr., “Overcast: Reliable multicasting with an overlay network,” in 5th Symposium on Operating System Design and Implementation (OSDI), Dec. 2000.

[6] D. Pendarakis, S. Shi, D. Verma, and M.Waldvogel, “ALMI: An application level multicast infrastructure,” in 3rd Usenix Symposium on Internet Technologies and Systems (USITS),Mar. 2001.

[7] S. Shi, J. Turner, and M.Waldvogel, “Dimension Server Access Bandwidth and Multicast Routing in Overlay Networks”. In11th International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV’01), June2001.

[8] Michael R. Garey, David S. Johnson　Computers and intractability, A Guide to the Theory of NP-Completeness, W.H.Freeman and Company,1979.

[9]Yan Zhu, Min-You Wu, and Wei Shu, “Comparison Study and Evaluation of Overlay Multicast Networks” http://www.eece.unm.edu/~wu/course/540s03/reading.html

[10]Luc Devroye，Enumeration and Representation of Free Trees，in Data Structures and Algorithms，April 19, 2000 at http://jeff.cs.mcgill.ca/~luc/251.html.

[11] S. Shi and J. Turner. “Routing in Overlay Multicast Networks”. In Proc. of IEEE INFOCOM’02, June 2002.