International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 4 • 2017

Genetic K-Means Clustering for Route Discovery in Multi Cast Networks

Sakshi Arora and *Sunanda

Department of Computer Science & Engineering, Shri Mata Vaishno Devi University, Katra, Jammu and Kashmir
*E-mail: sakshi@smvdu.ac.in; sunanda.gupta@smvdu.ac.in (*corresponding author)*

Abstract: Determining of the core is an issue that has been the focus of research in optimum route discovery in multicast networks, as the choice of centre greatly influences the topology of network and hence the quality of service in such networks. The effectiveness of such a strategy may be evaluated in terms of delay and bandwidth, as a core-based tree is a shortest-path tree rooted at some centre node. An important question in research has been to determine the computational complexity of the core selection strategy where on the one hand the core may simply be chosen from some preselected set of nodes or some complex heuristics may be employed to chose a suitable core for dynamically changing topology such as in mobile ad hoc networks. In this paper, we propose an adaptive k-means clustering heuristic for core selection in multicast networks. The simulation results show that the proposed clustering strategy performs better than the existing algorithms in terms of delay variation subject to the end-to-end delay bound. The mathematical time complexity and the execution time of the proposed algorithm are comparable to those of the existing algorithms particularly for large Euclidian instances.

General Terms: Clustering, Multicast Networks.

Keywords: K-Means Clustering, Route Discovery, Genetic Algorithms.

1. INTRODUCTION

With the advent of internet and now the more demanding internet of things and service clouds offering service to multiple clients from large datacenters, there is a need for multicast transmission capability from one or more sources to multiple destinations. Relatively uncomplicated approaches suffice for multicast transmission in local area networks as compared to inter domain multicasting which becomes more challenging in presence of diverse network characteristics such as instance size and network topology.

Determining of the core in such networks is an issue that has been the focus of research recently, as the choice of centre greatly inf luences the shape of the network and therefore it's performance. Spanning trees and particularly Steiner trees, bounded diameter minimum spanning trees (BDMST) play a significant role in multicast routing. An optimal spanning of a large multicast group is expensive, hence majority of the route-discovery protocols in multicast networks use a core-based strategy for transmission from all the sources, such as core-based trees [1] and Protocol-independent Multicast [2],[3]. A core-based tree is a shortest-path tree rooted at some core node[4]. In this paper, a core-based BDMST has been used to model a suitably dynamic Multicast Network as this can reduce the router shortage and the overhead generated due to control messages. Further, core-based trees may also sometimes offer improved bandwidth utilization when compared to shortest trees [4].

In the light of the above discussion, it is pertinent to point out the relationship between the core selection methodology and the performance of route-discovery protocol. For effective performance in a multicast network, the quality of Service (QoS) parameters such as end-to-end delay, delay variation, loss, cost, throughput are to be improved. [5]. Network resources on the route may not confirm to the required QoS. For efficient multicast communication, it is required that the tree constructed satisfies the resource requirements [6].

A Steiner trees or a BDMSTs for a mobile ad-hoc networks are built using either the Source based Algorithms (SBA) or the Core Based Algorithms (CBA). In a SBA the source node connects with all the recipients, whereas in CBA, a centre or core is selected that acts as a root of the tree. Such a core-based tree is the shortest path tree root at the core node [7]. Thus, the strategy by which the core is selected in such networks plays a decisive role in ensuring optimum network performance.

The issues discussed in this paper include: Optimal Core Selection Algorithm that can minimize the average distance (delay) from the selected core to each receiver in a destination group and a Tree Generation Algorithm that can minimize the number of links used and at the same time to achieve the shortest end-to-end average delay.

2. LITERATURE REVIEW

It has been established well in literature that the selection of core directly affects the performance of multicast networks [9],[10],[11]Therefore, it is important to select a good core to have effective multicast. The center in core-based tree (CBT) is called as core [12] or Rendezvous Point (RP) in PIM-SM [13]. The core selection problem was first proposed by Wall [4]. A poor selection of core may lead to decline in QoS parameters such as cost, delay and congestion [15]. However, the core selection has been shown to be NP-complete problem[12,13,16], and therefore can be solved effectively using heuristics. Several heuristic approaches to this problem have been discussed in literature [17], [18].

Several studies have focused on constructing core-based multicast tree satisfying delay and delay variation constraints. The issue of minimizing multicast delay variation under the end-to-end delay constraints is discussed as the delay and delay variation bounded multicast tree (DVBMT) problem [19]. In DVMA [19], it is assumed that the complete topology is maintained at each node. It constructs a spanning tree satisfying the delay constraint only. It then searches through the candidate paths satisfying the delay and delay-variation constraints.. The drawback of this approach is that, if the k-shortest paths to the destination group, violate a delay constraint, then a longer path may have to be searched. Also, the time complexity of DVMA is prohibitive.

Another approach is delay and delay variation constrained algorithm (DDVCA)[20], It first calculates the delay of the least delay paths from each destination node to all the nodes. The node that has the minimum delay-variation is chosen as the core node. The source node sends a single packet to the core node. The core node then forwards the message to all the receivers through the minimum delay path. The time complexity of DDVCA is $O(mn^2)$ where m is the number of destination nodes and n is the total number of nodes in the network.

AKBC (Ahn, Kim, Bang, Choo) algorithm [21] and AKC (Ahn, Kim, Choo) algorithm [22] have also been proposed in the literature. Both AKBC and AKC produce a core-based multicast tree under delay and delay-variation constraints. Potential core nodes are identified that have the same multicast delay-variation for each destination node. From amongst them the node with the minimum delay-variation is selected as the core.

However, all these algorithms have been studied on adirected graphs only. Ahn, Kim and Choo[15] proposed an algorithm that constructs a multicast tree with low delay-variation in a realistic

The framework proposed in [24] for networked multimedia system is an adaptive framework which changes dynamically based on user demands and preferences. It has been demonstrated that non uniform environments both hardware and software make interoperability problem challenging. A distributed core selection and migration protocols for mobile ad hoc networks [25] has also been suggested, in which the median node of the multicast tree is used to determine the core instead of the median of the entire network. Median of the tree has been used in place centroid. The cost of this multicast tree is then computed as the sum of weights of all the edges in the tree. This gives the total bandwidth required for transmitting a message from a source to multiple destinations. DSDMR [26] is an adaptive core selection protocol that uses an adaptive two-direction join mechanism. DSDMR focuses on poor scalability resulting from high control overhead, vulnerability of the multicast core and unnecessary delays. DSDMR has been shown to create low cost tree close to optimal greedy strategy with very low control overhead and delays. A Tabu search TS [28] based core selection algorithm, TRPSA [27] has been proposed that generates a local solution after a finite number of iterations by using memory lists. The best local solution is archived so that the research process does not get stuck in a loop. However, the method suffers from loosely defined termination criterion and an empirically determined size of the Tabu list both of which a strong bearing on the performance of the algorithm.

Author in [29] has proposed a core selection strategy DRGH which yields promising results but the selection of core may not be consistently high quality due to the instance dependent nature of DRGH.

In this study it is proposed to implement Genetic Optimization techniques to improve the QOS particularly end-end delay. The performance of the proposed algorithm is evaluated through simulations. It is observed the algorithm outperforms the existing algorithms. The rest of the paper is organized as follows. The mathematical model developed to model a computer network is presented in Section 2. The proposed algorithm and its principle of working followed by the case studies and complexity analysis are presented in Section 3. The simulation results of our algorithm are presented in Section 4. Finally, we present the conclusion in section 5.

3. PROPOSED STRATEGY

Building on the fact that the shortest-path trees such as MSTs can minimize the delay in Multicast routing, an MST based approach to Multicast routing has been adopted. For core selection, k-means clustering is applied as it tends to minimize the intra-cluster variation measures. But the problem with k-means clustering is that it does not guarantee to converge to global minima. Since Genetic Algorithm (GA) is better at avoiding entrapment in local optima, the k-means clustering has been enhanced using GA. Agglomorative approach has been used in which the points are first clustered using the k-means clustering and then fed as input to the GA for improvement. MST is then constructed for the intra-cluster cores again using genetic algorithm. The number of clusters k has to be given in advance. The advantage of using MST- based algorithm is that it does not assume that data points are grouped around centers or separated by regular geometric curve. Thus the shape of the cluster boundary does not have a significant impact on the performance of the algorithm.

3.1. Problem Model

The network scenarios have been modeled using eucledian graphs, in which vertices represent nodes or switches and the edges represent the links between them. The graphs have been generated to be consisting of k different neighborhoods; each neighborhood is delineated from the others by means of locality characteristics. A framework of a Eucledian graph is constructed consisting of k nodes. Each node of this graph is replaced with a neighborhood of 15 nodes.

3.2. Clustering Using MST for Route Discovery

Clustering has been considered for route optimality in the present study because clustering ensures minimum intra-cluster distance [30]. Cost of clustering can be minimized if minimal spanning tree (MST) is used for clustering, as MST ignores many possible links between the clustered points. Further, the MST based clustering algorithms are capable of detecting clusters with various shapes and size [30]. Unlike the non-MST based clustering algorithms, the MST clustering algorithm do not require spherical shapes structure of the underlying nodes. The EMST clustering algorithm [31] uses the Euclidean minimum spanning tree of a graph to produce the structure of point clusters in the n-dimensional Euclidean space [29],[31].

The proposed strategy starts with disjoint clustering, which places each of the n objects in an individual cluster [32]. The neighborhood is analyzed and the clusters which have a high similarity quotient SQ are merged. SQ is defined as the measure of compactness of the cluster.. In the second iteration, the next level merger of clusters is done and the algorithm iterates till the number of clusters less than equal to k. In order to measure the efficacy of clustering, a measure based upon the radius and diameter of each subtree (cluster) is used. The radius and diameter values of each cluster are expected low value for good cluster. If these values are large the points (objects) are spread widely and may overlap. The cluster tightness measure is within a threshold of the cluster estimate of clustering effectiveness, however it is possible to devise inter-cluster measure also, to better measure the separation between the various clusters. The Cluster compactness measure is based on the variance of the data points distributed in the subtrees (clusters).

3.3. Genetic Algorithm for MST

The Genetic Algorithm (GA) is a population-based heuristic, which optimizes by simulating the process of natural evolution. A population of potential solution to the problem are generated. Fitness functions (in this case the delay and delay variance) are used to evaluate each solution. Operators such as selection, reproduction, crossover and mutation then act on these population strings to yield better solutions during successive iterations. Algorithm terminates till an appropriate solution is reached or a specific number of generations have been obtained.

A typical GA is as follows

Start.

1: Generation t=0;

2: Initialize the population P(t);

3: While not termination criteria do

(i) Evaluate the population P(t);

(ii) Apply crossover & mutation to P(t) to yield P'(t);

(iii) Evaluate P'(t);

(iv) P(t+1) = select from P'(t);

(v) t=t+1;

End while.

A chromosome is represented by a string of length N, where N represents the number of clusters or cores in the graph. Each element r of the chromosome represents a path between source s and destination n. The duplicate strings in the population are discarded. Delay metric is used to evaluate the individuals of P. A roulette selection operator is then applied over P to generate the next evolutionary population P. crossover operator over a selected pair of individuals is applied and some genes in each chromosome of the new population are randomly changed (mutated), obtaining a new solution. The process continues until a stop criterion or the given maximum number of generations, is satisfied.

3.4. Algorithm

1. Determine the number of clusters k.

2. Initialize k cluster centres or cores based on mode.
3. For each node, compute the core nearest to it by using hamming distance and assign the node to this cluster.
4. Re-compute the new cluster center Q (mean of data points in cluster)
5. Stop when cluster centres do not improve.
6. For each cluster do
   (a) Select s nodes nearest to all the nodes using the all-pair shortest path algorithm.
   (b) For each s construct MST rooted at this centre using GA
   (c) Populate the Master GA with the above MSTs
   (d) Apply the crossover operator between the randomly chosen MST and the MST constructed using Q as root.
   (e) Perform mutation and get the new population.
   (f) Repeat till termination criteria.
   (g) Choose the root of the least cost MST as the Cluster centre or core.
7. Apply GA to construct MST using the cluster centres as nodes.
8. The least cost MST is the required optimal route.

4. RESULTS & DISCUSSIONS

The simulation program is written in C and runs in a SUN SPARC-20 workstation. The network simulated is devised with a 4-dimensional hypercube. The bandwidth of each link is uniform as 10Mbps. For each simulation run, 10000 packets have been randomly generated as a Poisson process and the average size of each packet is between 20-30 Kb.

The parameter settings for genetic algorithm have been as shown in table 1

Table 1
Parameter settings for genetic algorithm

Length of chromosomes	n
Number of chromosomes	3n
Number of Generations	5n
Crossover Rate	0.8
Mutation Rate	0.5
Selection Method	Tournament selection

Fig. 1 shows the simulation results of multicast delay-variations versus the number of nodes on a network, where delay variation is the difference between the first time of the reception of a multicast packet by a receiver of the multicast group and the last reception of the same multicast packet by another receiver of the multicast group. The destination nodes constitute 15% of the overall network nodes and 5% of the nodes are sources. The delay-variation of the proposed scheme is found to be better than that of DDVCA and DRGH. Since our algorithm selects the best path out of k-shortest paths from source to the best core node. And it can be seen that our algorithm incurs much less delay, particularly in the cases of heavy loading of multicast traffic in which every node is a multicast message source.

Fig. 2 demonstrates that the average delay incurred with various group sizes. From the figure, it is clear that our algorithm is scalable with the randomly generated member distributions. It can be easily observed that the execution time of our proposed algorithm is comparable to DVCA and DRGH. However, the proposed algorithm with k-shortest paths has more execution time than the other existing algorithms. This is because the execution time of our proposed algorithm with k-shortest paths depends on the value of k. It can be observed that the rate of increase of k goes down as the number of nodes increases. Therefore, the rate of increase in execution time also goes down with the increase in the number of nodes.

To establish the effectively of the proposed approach the results of the proposed algorithm have also been compared with the standard K-means Clustering algorithm and the standard genetic algorithm (SGA) based route discovery. The results tabulated in table 2 shows the average computational times (in sec) of the three strategies. Simple k-means algorithm takes the most time when applied on eucledian graphs, particularly when the number of nodes is large. Genetic Algorithm performs average as the increase in time complexity is not exponential. The best results can be seen in case of the proposed algorithm where the time complexity scales well with the increase in the number of nodes.

Table 2
Table captions should be placed above the table

Number of Nodes	*k-means clustering*	*GA based clustering*	*GA based k-means clustering*
50	7.67	4.76	3.42
100	129.3	30.55	4.65
200	759.25	202.37	5.62
500	1343.81	783.66	7.11
1000	> 15000	1024.30	31.40

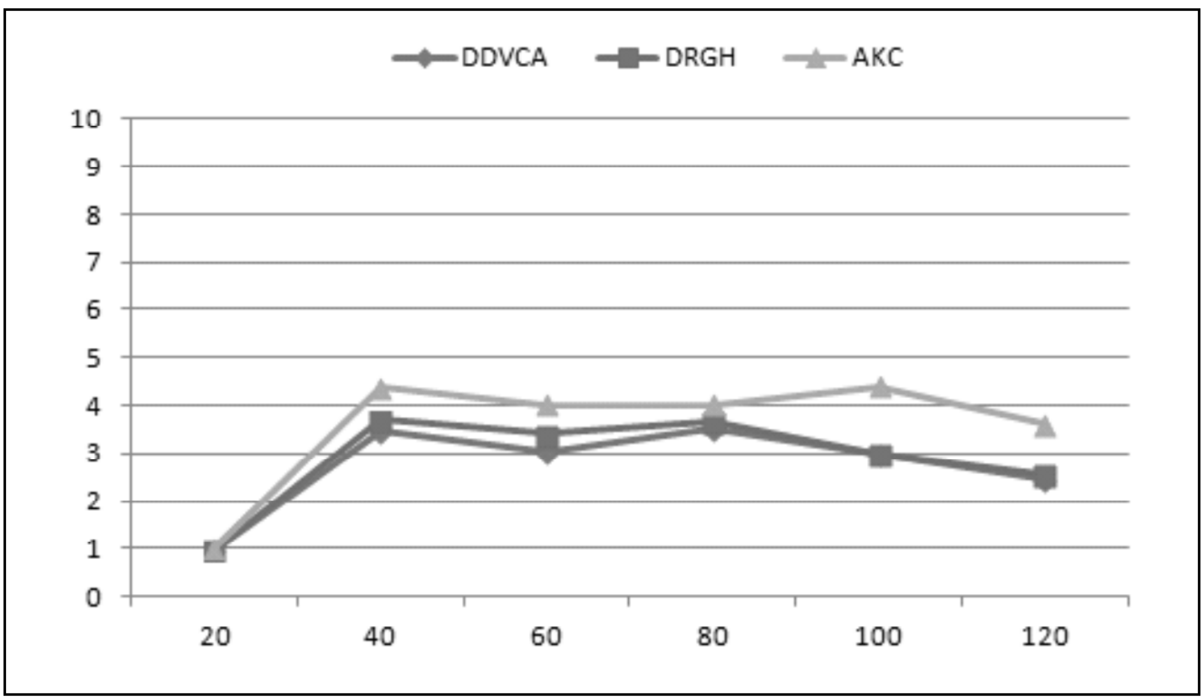

Figure 1: Comparison of multicast delay-variations for varying Network sizes

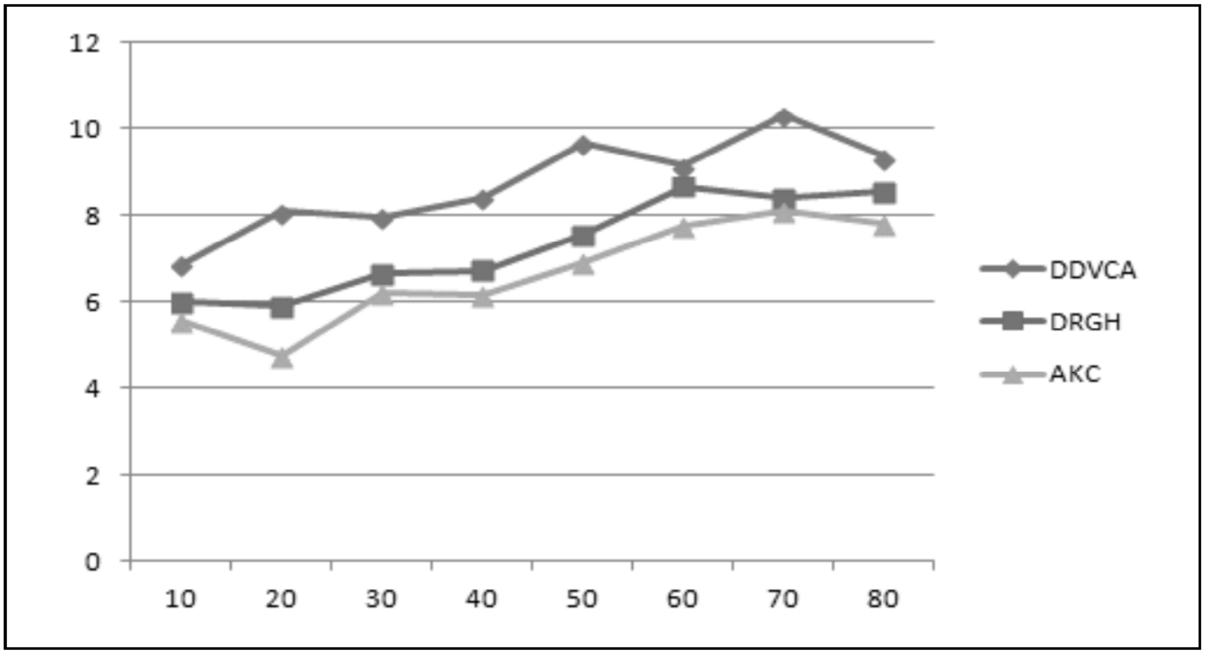

Figure 2: Comparison of multicast delay-variations for varying group sizes

5. CONCLUSION

In this paper, we have proposed a new framework to improve the core selection in clusters using k-means clustering and subsequent improvement of the cluster using genetic algorithm. The proposed algorithm is tested in multicast networks for route discovery and show that refined initial starting points lead to improved solutions. The method is scalable and consistent for various instances of graph problems. Experimental results show that the proposed algorithm achieves better results than the existing algorithm when applied to Euclidean data sets. Our future work is focused on extending this algorithm to support multiple QoS criteria imposed by Mobile receivers across the network.head.

REFERENCES

[1] Das, Subir Kumar, B.S.Manoj, and C. Siva Ram Murthy. "A dynamic core based multicast routing protocol for ad hoc wireless networks". Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing. ACM. 202.

[2] Deering, Stephen, *et al.* "Protocol independent multicast (pim): Motivation and architecture." *Work in Progress* (1995).

[3] Deering, Stephen, *et al.* Protocol independent multicast (PIM), sparse mode protocol: Specification. Internet Draft, March, 1994.

[4] Wang, Ting-Yuan, Lih-Chyau Wuu, and Shing-Tsaan Huang. "A scalable core migration protocol for dynamic multicast tree." *Parallel and Distributed Systems, 2002. Proceedings. Ninth International Conference on*. IEEE, 2002.

[5] Chen, S. and K. Nahrsted, 1998. An overview of quality of service routing for next-generation high-speed networks: Problems and solutions. IEEE Netw., 12: 64-79. DOI: 10.1109/65.752646.

[6] Wang, Bin, and Jennifer C. Hou. "Multicast routing and its QoS extension: problems, algorithms, and protocols." *Network, IEEE* 14.1 (2000): 22-36.

[7] Chandrasekar, V., and K. Baskaran. "End to End Delay Improvement in Heterogeneous Multicast Network using Genetic Optimization." *Journal of Computer Science* 8.9 (2012): 1514.

[8] Van Hentenryck, P., & Michel, L. Control abstractions for local search. In *Principles and Practice of Constraint Programming–CP 2003* (pp. 65-80).

[9] Calvert, Kenneth L., Ellen W. Zegura, and Michael J. Donahoo. "Core selection methods for multicast routing." *Computer Communications and Networks, 1995. Proceedings., Fourth International Conference on*. IEEE, 1995..

[10] Fleury, Eric, Yih Huang, and Philip K. McKinley. "On the performance and feasibility of multicast core selection heuristics." *Computer Communications and Networks, 1998. Proceedings. 7th International Conference on*. IEEE, 1998.

[11] Thaler, David G., and Chinya V. Ravishankar. "Distributed center-location algorithms." *Selected Areas in Communications, IEEE Journal on* 15.3 (1997): 291-303.

[12] Ballardie, A. "Core Based Trees (CBT version 2) Multicast Routing—Protocol Specification—." (1997).

[13] Fenner B. Protocol independent multicast-sparse mode (PIMSM): Protocol specification (revised). Internet RFC 4601, August 2008.

[14] Wall, David Wayne. "Mechanisms for broadcast and selective broadcast." (1980).

[15] Kabat, Manas Ranjan, Manoj Kumar Patel, and Chita Ranjan Tripathy. "A heuristic algorithm for core selection in multicast routing." *Journal of Computer Science and Technology* 26.6 (2011): 954-961.

[16] Oliveira, Carlos AS, and Panos M. Pardalos. "A survey of combinatorial optimization problems in multicast routing." *Computers & Operations Research* 32.8 (2005): 1953-1981.

[17] Chung, Sung Mo, and Chan-Hyun Youn. "Core selection algorithm for multicast routing under multiple QoS constraints." *Electronics Letters* 36.4 (2000): 1.

[18] Moonseong, K. I. M., et al. "On efficient core selection for reducing multicast delay variation under delay constraints." *IEICE transactions on communications* 89.9 (2006): 2385-2393.

[19] Zhengying, Wang, Shi Bingxin, and Zhao Erdun. "Bandwidth-delay-constrained least-cost multicast routing based on heuristic genetic algorithm." *Computer communications* 24.7 (2001): 685-692.

[20] Sheu, Pi-Rong, and Shan-Tai Chen. "A fast and efficient heuristic algorithm for the delay-and delay variation-bounded multicast tree problem." *Computer Communications* 25.8 (2002): 825-833.

[21] Ahn, Youngjin, et al. "On algorithm for the delay-and delay variation-bounded multicast trees based on estimation." *High Performance Computing and Communications*. Springer Berlin Heidelberg, 2005. 277-282.

[22] Ahn, Soobeen, Moonseong Kim, and Hyunseung Choo. "Efficient algorithm for reducing delay variation on delay-bounded multicast trees in heterogeneous networks." *Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE*. IEEE, 2008.

[23] Zahn, Charles T. "Graph-theoretical methods for detecting and describing gestalt clusters." Computers, IEEE Transactions on 100.1 (1971): 68-86.

[24] Gómez-Montalvo, Jorge, Myriam Lamolle, and Ernesto Exposito. "A Multimedia Ontology Driven Architecture framework (MODA) for networked multimedia systems." Networked Digital Technologies, 2009. NDT'09. First International Conference on. IEEE, 2009.

[25] Gupta, Sandeep K. S., and Pradip K. Srimani. "Adaptive core selection and migration method for multicast routing in mobile ad hoc networks." Parallel and Distributed Systems, IEEE Transactions on 14.1 (2003): 27-38.

[26] Tseng, Chih Jen, and Chyou Hwa Chen. "The performance of QoS aware IP multicast routing protocols." Networks 42.2 (2003): 97-108.

[27] Sahoo, Satya Prakash, Manas Ranjan Kabat, and Ashis Kumar Sahoo. "Tabu search algorithm for core selection in multicast routing."Communication Systems and Network Technologies (CSNT), 2011 International Conference on. IEEE, 2011.

[28] Glover, Fred. "Tabu search—part II." ORSA Journal on computing 2.1 (1990): 4-32.

[29] Arora, S., and Gupta S., "Effect of core selection strategy in determining optimal BDMST in Multicast Networks". Advances in Computer Science and Information Technology (ACSIT) Volume 2, Number 3; (2015) pp. 249-251.

[30] Zahn, Charles T. "Graph-theoretical methods for detecting and describing gestalt clusters." Computers, IEEE Transactions on 100.1 (1971): 68-86.

[31] Preparata, F. "P. and Shamos, M." I., Computational Geometry: An introduction, Springer-Verlag, New York (1985).

[32] Jain, Anil K., and Richard C. Dubes. Algorithms for clustering data. Prentice-Hall, Inc., 1988. IJCA™ : *www.ijcaonline.org*