2012 Fourth International Conference on Computational Intelligence, Communication Systems and Networks

Minimizing Network Coding Nodes in Multicast Tree Construction via Genetic Algorithm

Shee Eng Tan, Zhan Wei Siew, Yit Kwong Chin, Scott Carr Ken Lye, Kenneth Tze Kin Teo
Modelling, Simulation & Computing Laboratory, Minerals & Materials Research Unit
School of Engineering and Information Technology
Universiti Malaysia Sabah
Kota Kinabalu, Malaysia
msclab@ums.edu.my, ktkteo@ieee.org

***Abstract*— Network coding is a method that increases network throughput by encoding several packets with single packet size length and forwards the packet in a single transmission time slot. At the same time, network coding increases the complexity of packets management and delay of network due to the waiting time for network coding opportunity. A solution based on improved genetic algorithm is proposed to optimize the network coding node resources in network coding. Genetic algorithm will search a multicast tree that fulfils the desired throughput with a desired multicast rate. Mutation rate of the genetic algorithm will change based on the previous solution to avoid from being stuck on the local optima. The simulation result shows that with given multicast rate, improved genetic algorithm is able to search and construct multicast tree with minimal usage of network coding nodes.**

Keywords-network coding; genetic algorithm; optimization; multicast.

I. INTRODUCTION

Network coding [1, 2] is a method that increases network throughput by combining packets from different flows and forward with less transmission. The transmission with one packet size length contains more than one packet. Unlike traditional store and forward method, which carry one packet per transmission. Network coding are widely used in wireless mesh network [3], wired network, wireless sensor network [4], and vehicular ad hoc network due the improvement on throughput are significant.

Fig. 1 shows the comparison of traditional store and forward and network coding to illustrate the advantage of network coding. Node S is source of the network and nodes *T1*, *T2* represent destination or sink. Node S cannot directly contact with node *T1* and *T2*, the information will route through the intermediate node 1, 2, 3, and 4. Packet *a* and packet *b* are sending from node *S* to *T1* and *T2*. The route of information flow of packet a for traditional store and forward method are *S* to 1, 1 to *T1*, *S* to 2, 2 to *T2*, in total this require 4 transmission time slot, packet b repeat the same procedure like packet *a*, therefore the total transmission time slot require are 8 time slot. The transmission time slot require for network coding method are less than traditional method, 2 transmission time slot is used to transfer packet *a*

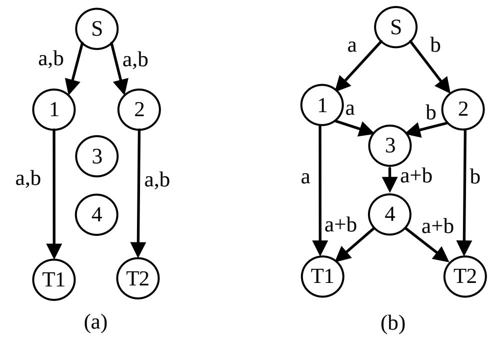

Figure 1. Sample network: (a) Traditional store and forward. (b) Network Coding.

and packet *b* are send to node 1 and node 2 as shown at Fig. 1, node 1 will multicast packet a to node 3 and T1 and node 2 will multicast packet *b* to node 3 and T1, in this stage the total transmission time slot used are 4 time slot, node 3 will perform network coding by combine packet *a* and packet *b* into single packet with same size length by XOR, and forward to node 4, node 4 will multicast to *T1* and *T2* and since *T1* and *T2* got another packet, *T1* and *T2* can decode the combine packet to get another packet with XOR. In total network coding only require 6 transmission time slots to distribute both packet to both destination.

Traditional network coding method will perform network coding at all possible intermediate node to increase the throughput for overall network [5], but network coding also increase the complexity of computational at the intermediate node and increase the usage of buffer [6] use to store packet for decode incoming packet. Research on opportunities scheduling of network coding prove that waiting time for the packet will increase the coding chance in intermediate node [7], but network coding also brings delay for the network due to the waiting time [8, 9]. Therefore, by fully brings the benefit of network coding and minimize the overhead of all kinds for network coding has become an importance optimization problem for network coding problem [10].

Optimization problem in network coding means at a given topology, minimize all kinds of overhead of network coding and achieve the multicast rate needs for network coding. Kim [11, 12] proven that optimizing of coding

978-0-7695-4821-0/12 $26.00 © 2012 IEEE
DOI 10.1109/CICSyN.2012.79

resources are NP-hard problem, and propose simple genetic algorithm approach to solve this problem. In this paper, genetic algorithm with adaptive mutation rate is proposed to overcome the local optima problem. Proposed genetic algorithm will construct multicast tree base on the given topology and multicast rate, and reduce the number of the network coding resource [13].

II. GENETIC ALGORITHM

Genetic algorithm is an optimal solution searching method that is based on Darwin's biological evolution process. Theoretical research of Genetic algorithm are improving and mature enough to solve many optimization problems such as batch process control and traffic flow control [14, 15]. Fig. 2 shows the flow chart of genetic algorithm. Genetic algorithm is a specialized method can search for a global optimization solution. The general idea is to initiate population and put this population in some environment. In genetic algorithm, the environment referred to as fitness function. Fitness function will evaluate the fitness of each chromosome and least fit chromosomes will be eliminated. Finally the populations that can survive in given fitness function are the optimal solution.

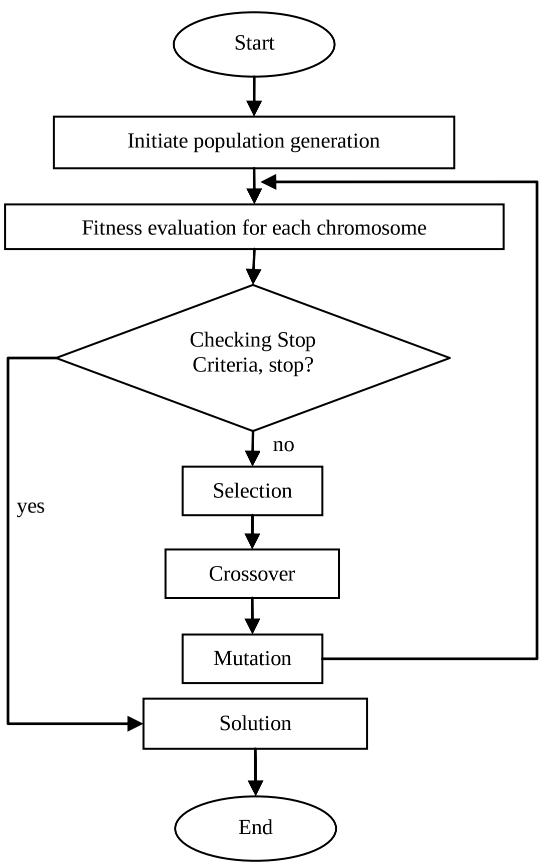

Figure 2. Flow chart of genetic algorithm.

III. NETWORK CODING OPTIMIZATION MODEL

A. Chromosome Defination

Binary encoding is a simple and easy to computational of the biological evolution process such as crossover and mutation operation. The use of binary symbol is to represent the individual's existent. In optimization of network coding, the chromosome represents the existence of the link of the node. Every node in a given topology with more than 1 incoming flow will potentially become a network coding node. Let say in a given topology there are *i* nodes that has network coding node potential. A chromosome will split into *i* segments to represent the link state of each potential node. Let say a potential node *i* has *j* incoming flow and k outgoing flow, the total bit of segment i is shown as in (1).

$$g_i = j \times k \quad (1)$$

Let's assume *i* potential coding is present in a given topology, V_i represents the i^{th} potential coding node, $V = \{V_1, V_2, V_3 \ldots V_i\}$. The total bits needed for a complete chromosome is as (2).

$$n = \sum_{V_1}^{V_i} g_i \quad (2)$$

Fig. 3 shows an example of a potential node V_2 with 3 incoming links and 2 outgoing links, since V_2 have 3 incoming links, g_2 have 3 bit to represent input link x_1, x_2, x_3 to link y_1 and another 3 bit to represent input link x_1, x_2, x_3 to link y_2, in total g_2 have 6 bits. Example as in Fig. 3, a_1=000 represent that link y_1 is not going to send any packet, a_2=101 mean the node will perform network coding by combine x_1 and x_3, and send through link y_2. Other nodes at the network with single incoming flow will just forward the packet, so chromosome only contains link state for node with several incoming flow or potential coding node.

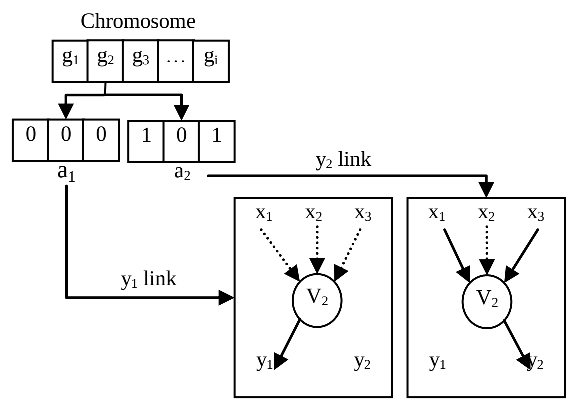

Figure 3. Chromosome representative for node v with 3 incoming flow and 2 outgoing flow.

B. Population initialize

Population initiation in this paper is randomly generated to fill all 0 and 1 in the chromosomes. More population requires less generations to get the solution while less population needs more generations to get to the solution. More population will lead to more precise solution but requires more computational time.

C. Feasibility Test

For a given network, the node with several inputs are considered to be potential node that can perform network coding, while single incoming node are simple packet forwarding. A chromosome is broke into several segments base on how many potential coding nodes, and the chromosome will generate random bits of 0 and 1. The feasibility test on a new generated chromosome is conducted. If a new generate chromosome have an empty link on the middle of the route path, this chromosome can be conclude to be invalid. Fig. 4 show an example topology, feasibility test is conducted on chromosomes.

Information flow equations are construct base on the topology of Fig. 4. S node will send 3 packets x_1, x_2 and x_3 to T_1, the equation are generating base on the given topology and link. $\beta_{ei,ej}$ show that the packet are flow from link e_i to link e_j, $y(e_i)$ content the inflowing information for link e_i.

$$
\begin{aligned}
&y(e_1) = \beta_{1,e1}\,x(1) + \beta_{2,e1}\,x(2) + \beta_{3,e1}\,x(2)\\
&y(e_2) = \beta_{1,e2}\,x(1) + \beta_{2,e2}\,x(2) + \beta_{3,e2}\,x(2)\\
&y(e_3) = \beta_{1,e3}\,x(1) + \beta_{2,e3}\,x(2) + \beta_{3,e2}\,x(2)\\
&y(e_4) = \beta_{e1,e4}\,y(e_1)\\
&y(e_5) = \beta_{e3,e5}\,y(e_3)\\
&y(e_6) = \beta_{e1,e6}\,y(e_1)\\
&y(e_7) = \beta_{e1,e7}\,y(e_1)\\
&y(e_8) = \beta_{e4,e8}\,y(e_4) + \beta_{e2,e8}\,y(e_2) + \beta_{e5,e8}\,y(e_5)\\
&y(e_9) = \beta_{e4,e9}\,y(e_4) + \beta_{e2,e9}\,y(e_2) + \beta_{e5,e9}\,y(e_5)\\
&y(e_{10}) = \beta_{e4,e10}\,y(e_4) + \beta_{e2,e10}\,y(e_2) + \beta_{e5,e10}\,y(e_5)\\
&y(e_{11}) = \beta_{e3,e11}\,y(e_3)\\
&y(e_{12}) = \beta_{e3,e12}\,y(e_3)\\
&y(e_{13}) = \beta_{e10,e13}\,y(e_{10}) + \beta_{e12,e13}\,y(e_{12})\\
&y(e_{14}) = \beta_{e6,e14}\,y(e_6) + \beta_{e8,e14}\,y(e_8)\\
&y(e_{15}) = \beta_{e6,e15}\,y(e_6) + \beta_{e8,e15}\,y(e_8)\\
&y(e_{16}) = \beta_{e14,e16}\,y(e_{14}) + \beta_{e7,e16}\,y(e_7) + \beta_{e9,e16}\,y(e_9) +\\
&\qquad \beta_{e11,e16}\,y(e_{11}) + \beta_{e13,e16}\,y(e_{13})\\
&y(e_{17}) = \beta_{e10,e17}\,y(e_{10}) + \beta_{e12,e17}\,y(e_{12})
\end{aligned}
$$

Example show in Fig. 4, sink is connected with e_{15}, e_{16}, and e_{17}, the complete equation for $y(e_{15})$, $y(e_{16})$, and $y(e_{17})$ are feasible test equation. Reorganize $y(e_{15})$, $y(e_{16})$ and $y(e_{17})$, the equation of packet x_1, x_2, and x_3 flow to sink T_1, are shown in Table I

$$
\begin{aligned}
&y(e_{15}) = \beta_{e1,e6}\beta_{e6,e15}y(e_1) + \beta_{e1,e4}\beta_{e4,e8}\beta_{e8,e15}y(e_1) +\\
&\qquad \beta_{e2,e8}\beta_{e8,e15}y(e_2) + \beta_{e3,e5}\beta_{e5,e8}\beta_{e8,e15}y(e_3)\\
&y(e_{16}) = \beta_{e1,e6}\beta_{e6,e14}\beta_{e14,e16}y(e_1) + \beta_{e1,e7}\beta_{e7,e16}y(e_1) +\\
&\qquad \beta_{e1,e4}\beta_{e4,e9}\beta_{e9,e16}y(e_1) + \beta_{e3,e11}\beta_{e11,e16}y(e_3)\\
&\beta_{e1,e4}\beta_{e4,e10}\beta_{e10,e13}\beta_{e13,e16}y(e_1) + \beta_{e2,e10}\,\beta_{e10,e13}\beta_{e13,e16}y(e_2) +\\
&\beta_{e3,e5}\,\beta_{e5,e10}\beta_{e10,e13}\beta_{e13,e16}y(e_3) + \beta_{e3,e12}\beta_{e12,e13}\beta_{e13,e16}y(e_3)\\
&y(e_{17}) = \beta_{e1,e4}\beta_{e4,e10}\beta_{e10,e17}y(e_1) + \beta_{e2,e10}\beta_{e10,e17}y(e_2) +\\
&\qquad \beta_{e3,e5}\beta_{e5,e10}\beta_{e10e17}y(e_3) + \beta_{e3,e12}\beta_{e12,e17}y(e_3)
\end{aligned}
$$

TABLE I. FEASIBILITY TEST TABLE

Packet	**Flow link**		
	e_{15}	***e_{16}***	***e_{17}***
X_1	$\beta_{e1,e6}\beta_{e6,e15}+$ $\beta_{e1,e4}\beta_{e4,e8}\beta_{e8,e15}$	$\beta_{e1,e6}\beta_{e6,e14}\beta_{e14,e16}+$ $\beta_{e1,e7}\beta_{e7,e16}+$ $\beta_{e1,e4}\beta_{e4,e9}\beta_{e9,e16}$	$\beta_{e1,e4}\beta_{e4,e10}\beta_{e10,e17}$
X_2	$\beta_{e2,e8}\beta_{e8,e15}$	$\beta_{e2,e10}\beta_{e10,e13}\beta_{e13,e16}$	$\beta_{e2,e10}\beta_{e10,e17}$
X_3	$\beta_{e3,e5}\beta_{e5,e8}\beta_{e8,e15}$	$\beta_{e3,e11}\beta_{e11,e16}+\beta_{e3,e5}$ $\beta_{e5,e10}\beta_{e10,e13}\beta_{e13,e16}+$ $\beta_{e3,e12}\beta_{e12,e13}\beta_{e13,e16}+$ $\beta_{e3,e12}\beta_{e12,e13}\beta_{e13,e16}$	$\beta_{e3,e5}\beta_{e5,e10}\beta_{e10e17}$ $+\ \beta_{e3,e12}\beta_{e12,e17}$

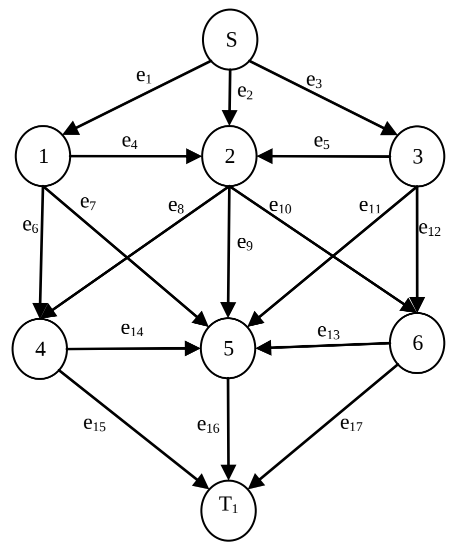

Figure 4. Simple network with 1 source and 1 sink.

Table I is feasible test for packet x_1, x_2, and x_3 to link e_{15}, e_{16}, and e_{17}. Random generate a chromosome for Fig. 4 network, these chromosome can put into feasible test table to test the chromosome is valid or not. If the chromosome is valid, the total value get at each row is at least 1. For example, let say a random generate chromosome are 101010 1010101010101010, the link that are not in the chromosome are assume to be 1 to guarantee the packet flow through. For x_1 to link e_{15}, feasible test are 1, row for x_1 can skip since already have 1 on that row. For row 2, packet x_2 to link e_{15}, e_{16}, e_{17} is unavailable because row 2 get 0 at every column, so this chromosome is invalid.

D. Fitness Function

Fitness function is a very importance in genetic algorithm, unsuitable fitness function will lead the solution to another direction, which causes the genetic algorithm, may not get the solution. In minimize resource of network coding, fitness function used shown as in (3).

$$F(R)=\begin{cases} n_{node},\ feasible=1 \\ n_{\max}+1,\ feasible=0 \end{cases} \tag{3}$$

n_{max} represent the quantity of all possible coding node in a given network, n_{node} is the quantity of coding node that in chromosome. Feasible test will test on the new generate chromosome, if the test pass, n_{node} will be fitness value for that combination of chromosome, if fail than n_{max}+1 will be the fitness value for it. Besides that, the chromosome should meet the multicast requirement. The chromosome which fails to do it will be eliminating. It is easy to see that all 1 in the chromosome are the maximum coding node in the network, mean all coding node are chooses, therefore n_{max} will never bigger than n_{node}.

E. Selection

Genetic algorithm selection methods are categories to fitness proportionate selection, stochastic universal sampling, tournament selection, reward based selection. The selection method used in this paper is fitness proportionate selection, also known as roulette wheel selection. Roulette wheel selection requires low complexity computation power and fast convergence in nature, the benefit of roulette wheel selection are suitable for the cases that need to get solution in short time. At the same time, the chromosome that are most fit in the given environment will be use to substitute to the chromosome that are most weak in the population to increase the probability of being choose. Fig. 5 shows the flow chart of the roulette wheel selection used in this paper.

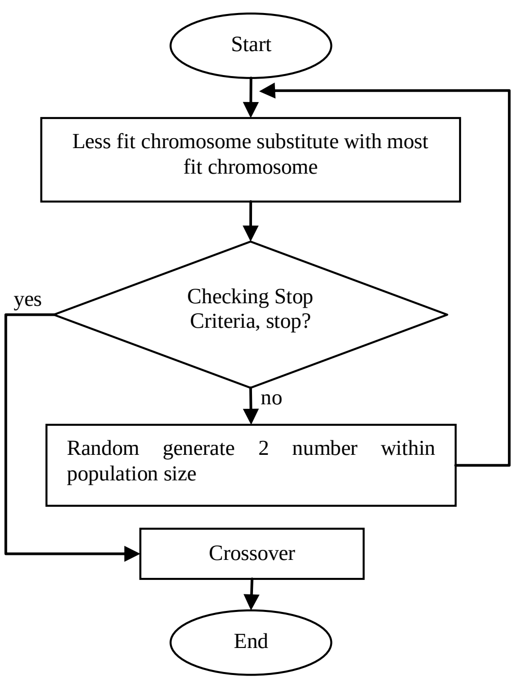

Figure 5. Selection operator.

F. Crossover

In biology evolution process, crossover plays an importance roll to reorganize the biology genetic. Crossover is genetic operator that increases the ability and speed of searching solution. Crossover between two chromosomes is not necessary success every time. Crossover rate are normally in between 0.5-0.9 base on the application. Fig. 6 shows the flow chart of the crossover operator.

Because of the chromosome definition in this paper is divided into each potential node and output link of potential node, therefore when crossover between 2 chromosomes, calculation will handle the chromosome block by block, only same potential node will crossover. Fig. 7 shows the example of crossover using on this paper.

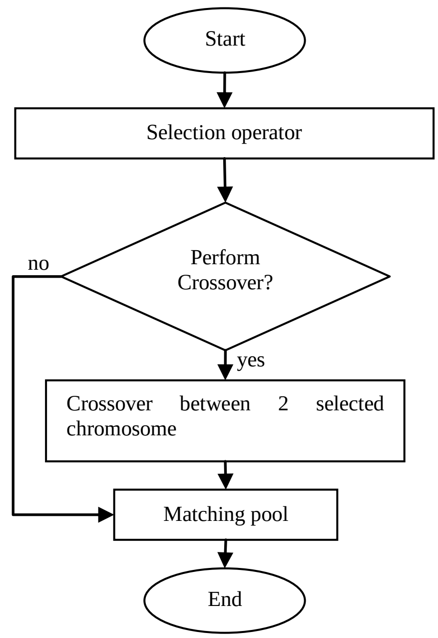

Figure 6. Crossover operator.

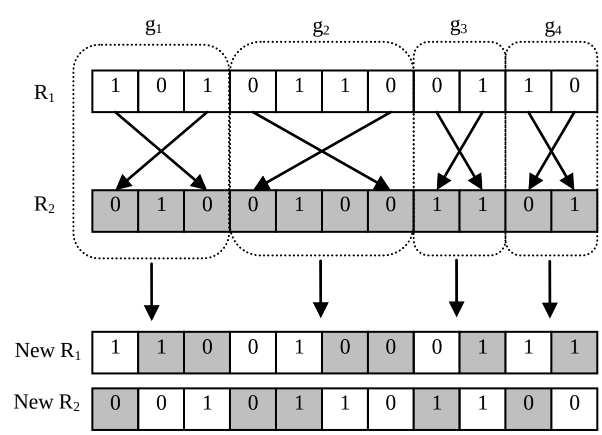

Figure 7. Crossover operation.

G. Mutation

In order to prevent the searching ability of genetic algorithm trap in local optimum, mutation are used to generate more possibility on the population to create more space to search the optimum solution. In this paper, the propose mutation rate will change according to the changes on the population on each generation. On the evolutional process, the similarity between populations will gradually increase after a certain generation, in this stage the solution will easy fall into local optimum solution. In this stage, add new chromosome into population is needed to improve the population quality. Proposed mutation rate are dynamic change base on the quantity of the same chromosome in the population. Besides, the same population on previous generation and current generation will affect the mutation rate. Mutation rates m are shown as in (4).

$$m = \min(\alpha, \exp(-\frac{(p-\varepsilon)}{p} \times 10), \Delta c) \qquad (4)$$

where m is mutation rate, α is the limit of the mutation rate, p is population size, ε is the quantity of same chromosome in population, Δc is the value of same population that appear at the previous generation until current generation. The value Δc will start from 0. Every time the optimum solution does not change, Δc will become 1. When optimum solution change, then Δc will become 0.

IV. Simulation and Performance Evaluation

The environment of network topology is simulated in MATLAB M-file. The topology and link of the network are shown in Fig. 8. Bidirectional links are not in the scope of this paper, so of backward links are eliminated from the topology. S is the source of the network, *S* will transmit 2 packets named packet *a* and packet *b* to destination node t_1, t_2, t_3, and t_4. Node 8, 9, and 10 will become potential network coding node. All outgoing links of potential coding node are e_{14}, e_{15}, and e_{26}, and therefore in this case n_{max} is 3. The topology used needs 6 bit chromosome for evolution calculation. Crossover rate are fixed to 0.7, and mutation rate follows (4) as previous section. Maximum generations are 100 round and population size is 10.

In this section, simulation result of difference mutation rate is test on the given topology. Population size are fix at 10 for all cases, stop criteria is 100 round of generation or solution reach 80% of total population, mutation rate are 0, 0.05, 0.1 and a case of dynamic mutation. Fig. 9 is the multicast tree generated by improve genetic algorithm, instead of using all 3 potential coding node, 1 coding node is used and this shows that transmissions of e_8 and e_{13} are saved. The simulation with mutation rate 0 shown at Fig. 10 will converge to the solution within 100 generation, but certain combination set of initial populations will trap solution in local optima. Some of the solution converge for mutation rate with 0.05 is shown at Fig. 11 but converge the solution using 100 round, but mutation rate 0.1 shown at Fig. 12 appear to show 100 generation time. Fig. 13 show that mutation rate varies from 0 to 0.1 provide better solution within 10 round, but some of solution need more generation to enable successful mutations to avoid solution trap on local optima.

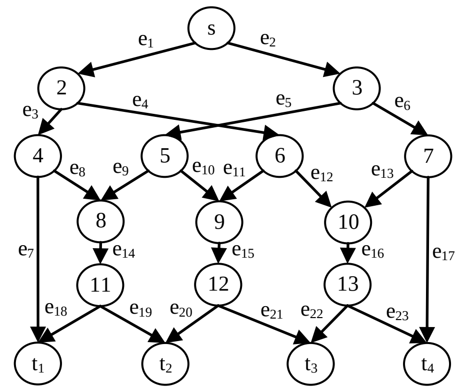

Figure 8. Simulation topology.

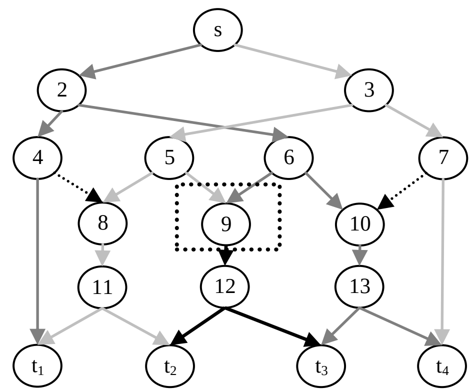

Figure 9. Potential flow by improve genetic algorithm.

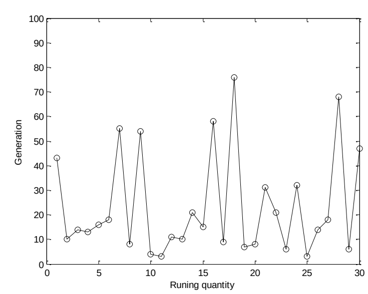

Figure 10. Generation needed for mutation rate 0.

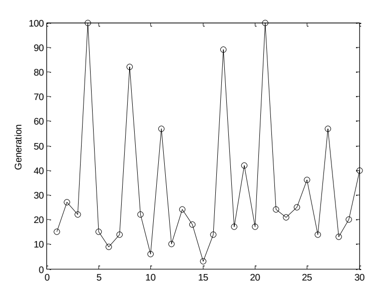

Figure 11. Generation needed for mutation rate 0.05.

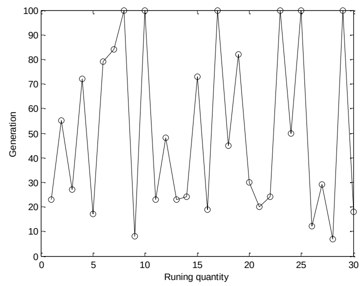

Figure 12. Generation needed for mutation rate 0.1.

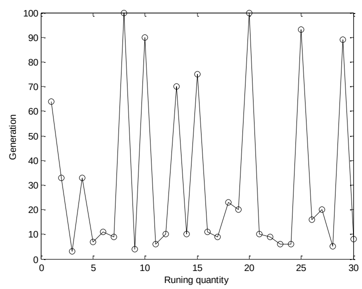

Figure 13. Generation needed for mutation rate 0 to 0.1.

V. CONCLUSION

Network coding is a proven method that increases network throughput. An increase in the usage of network coding nodes in the network will provide more throughputs. However, network coding also provide a side effect of delay and network cost. In this paper, genetic algorithm is used to minimize the usage of coding in the network in order to achieve optimization of advantage and disadvantage of network coding. Dynamic Mutation rate of the genetic algorithm will change base on the changes of the previous solution to avoid from falling into local optima. The simulation result shows that some solutions in dynamic mutation rate need more generations to converge because it manages to avoid being trap in local optima due to success of the improved genetic algorithm. For future work, multicast tree construction will consider packet flow for multiple sources to multiple sinks.

ACKNOWLEDGMENT

The authors would like to acknowledge the financial assistance of the Ministry of Higher Eduation of Malaysia (MoHE) under Exploratory Research Grant Schemes (ERGS), grant no. ERG0021-TK-1/2012, Fundamental Research Grant Schemes (FRGS), grant no. FRG0220-TK-1/2010 and scholarship support under MyMaster program.

REFERENCES

[1] R. Ahlswede, N. Cai, S.R. Li, and R.W. Yeung, "Network information flow," IEEE Transactions on Information Theory, vol. 46, no. 4, pp. 1204-1216, 2000.

[2] S.Y. Li, R. Yeung, and N. Cai, "Linear network coding," IEEE Transactions on Information Theory, vol. 49, no. 2, pp. 371-381, 2003.

[3] R.W. Yeung, Information Theory and Network Coding.USA: Springer, 2008.

[4] Z.W. Siew, A. Kiring, H.T. Yew, P. Neelakantan and K.T.K. Teo, "Energy Efficient Clustering Algorithm in Wireless Sensor Networks using Fuzzy Logic Control," Proc. 2011 IEEE Colloquium on Humanities, Science and Engineering Research (CHUSER 2011), 2011, pp.392-397, doi:10.1109/CHUSER.2011.6163758.

[5] B. Ni, N. Santhapuri, Z. Zhong, and S. Nelakuditi, "Routing with Opportunistically Coded Exchanges in Wireless Mesh Networks," 2nd IEEE Workshop on Wireless Mesh Networks 2006 (WiMesh 2006), 2006, pp. 157-159, doi: 10.1109/WIMESH.2006.288636.

[6] L. Zhang, S. Xiao, N. Cai, and J. Du, "An Opportunistic Network Coding Algorithm based on the Queue State and Network Topology," Proc. 2011 Seventh International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2011, pp. 290-293, doi: 10.1109/IIHMSP.2011.92.

[7] S.E. Tan, H.T. Yew, M.S. Arifianto, I. Saad, K.T.K. Teo, "Queue Management for Network Coding in Ad Hoc Networks," Proc. 2012 Third International Conference on Intelligent Systems, Modelling and Simulation (ISMS 2012), 2012, pp. 657-662, doi: 10.1109/ISMS.2012.113.

[8] J. Zhang, and P. Fan, "Optimal Scheduling for Network Coding: Delay v.s. Efficiency," Proc. 2010 IEEE Global Telecommunications Conference (GLOBECOM 2010), 2010, pp. 1-5, doi: 10.1109/GLOCOM.2010.5683947.

[9] Y.E. Sagduyu, and A. Ephremides, "Some Optimization Trade-offs in Wireless Network Coding," Proc. 2006 40th Annual Conference on Information Sciences and Systems, 2006, pp. 6-11, doi: 10.1109/CISS.2006.286422.

[10] Z. Ding, Z. Ma, and K.K. Leung, "Impact of Network Coding on System Delay for Multi-Source Multi-Destination Scenarios," Proc. 2010 IEEE International Conference on Communications (ICC 2010), 2010, pp. 1-5, doi: 10.1109/ICC.2010.5502760.

[11] M. Kim, M. Medard, V. Aggarwal, U.M. O'Reilly, W. Kim, C.W. Ahn, and M. Effros, "Evolutionary Approaches to Minimizing Network Coding Resources," Proc. 26th IEEE International Conference on Computer Communications (INFOCOM 2007), 2007, pp. 1991-1999, doi: 10.1109/INFCOM.2007.231.

[12] R. Koetter, and M. Medard, "An Algebraic Appraoch to Network Coding," IEEE/ACM Transactions on Networking, vol. 11, no. 5, 2003, pp. 782-795, doi: 10.1109/TNET.2003.818197.

[13] T. Ho, M. Medard, R. Koetter, D.R. Karger, M. Effros, J. Shi, and B. Leong, "A Random Linear Network Coding Approach to Multicast," IEEE Transactions on Information Theory, vol. 52, no. 10, 2006, pp. 4413-4430, doi: 10.1109/TIT.2006.881746.

[14] M.K. Tan, Y.K. Chin, H.J. Tham, and K.T.K. Teo, "Genetic Algorithm based PID Optimization in Batch Process Control," Proc. 2011 IEEE International Conference on Computer Applications and Industrial Electronics (ICCAIE 2011), 2011, pp. 162-167, doi: 10.1109/ICCAIE.2011.6162124.

[15] K.T.K. Teo, W.Y. Kow, and Y.K. Chin, "Optimization of Traffic Flow Within an Urban Traffic Light Intersection with Genetic Algorithm," Proc. 2010 Second International Conference on Computational Intelligence, Modelling and Simulation (CIMSiM 2010), 2010, pp. 172-177, doi: 10.1109/CIMSiM.2010.95.