A Study of Fluctuations in Genetic Algorithm Optimized Network in Data Centre

Okta Nurika[1], Mohd Fadzil Hassan[2], Nordin Zakaria[3], and Low Tan Jung[4]
{okta.rider@gmail.com[1], mfadzil_hassan@utp.edu.my[2],
nordinzakaria@utp.edu.my[3], lowtanjung@utp.edu.my[4]}

Department of Computer & Information Sciences, Universiti Teknologi PETRONAS,
Perak, Malaysia

Abstract. Study of fluctuation in genetic algorithm has been a sub-objective in genetic algorithm implementations. The reliability of genetic algorithm may vary based on implementation case, hence it is necessary to investigate its performance pattern for each implementation case. The purpose of this study is to observe the reliability of genetic algorithm in our previously simulated network optimization in a data centre. Previous researchers found fluctuation as random occurrence, mainly within small population. This paper's fluctuation observation revolves around our recent optimization of data centre's network. Our findings agree with the nature of genetic algorithm and other researches, where it is found that the fluctuation of fitness values in our case happened randomly in general, but it had higher probability with small population size. However, regardless of fluctuations that in average occurred during early stage of population generation, the near-optimal solutions with near maximum fitness values were able to be generated. This fact has proven the robustness of genetic algorithm itself.

Keywords: fluctuation; drift; genetic algorithm; network card optimization

1 Introduction

In this paper, we present the genetic algorithm (GA) fluctuation analysis of our previous work [1] in network optimization of a data centre in a simulated environment. Fluctuation in this context means the tendency of the individuals to move away from the near optimal solution, which is indicated by the degradation of fitness value. Paper [2] infers this incorrect convergence as 'drift'. However, different researchers have different perspectives of drift. For example, authors in [3] view genetic drift as a convergence to single solution regardless of its direction of fitness quality (higher or lower). Furthermore, it is concluded in [4] that GA convergence initially happens through selection, then later through genetic drift. This emphasizes that genetic drift can be beneficial. The above mentioned perceptions could bring a notion that GA fluctuation and genetic drift are actually the same thing, but in this case we clearly use 'fluctuation' term because of its single meaning, compared to the ambiguous 'drift'.

The importance of GA fluctuation study is to observe the behavior of GA in specific function. This is justified by [5] who observed genetic drift to forecast its own occurrence. The forecasting study of genetic drift, which in our case is called fluctuation was also done by [6] who proposed the finding of number of generations before genetic drift occurred in their water resources optimization. The next section will inform the simulated data centre's network optimization case that we use as a subject in this GA fluctuation study.

2 Optimization Case

The optimization subject that we observe is taken from [1] (the authors' previous work). It is a GA based network optimization on a data centre network, which consists of 47 servers, with every of them receiving different transmission characteristics with different combination of transmission size and packet size. The fitness value is each server's network card's processing speed (throughput) towards their particular received transmission. The maximum network card speed is 100 Mbps.

In order for the network card to achieve near-maximum receiving speed, it must be set according to the transmission characteristics that it receives, which include the transmission size and the packet size. The network card setting consists of 3 options, which are active wait, passive wait, and watermark mode as detailed in our paper [7] about the proposed mathematical models to represent the throughput generation of these 3 modes. The active wait triggers kernel interrupt to call the CPU to process every packet that comes. Furthermore, the passive wait utilizes polling based on timer option ranging from 10ms up to 200ms, with increment option every 10ms. Meanwhile, the watermark mode uses polling based on amount of packets captured (in Bytes unit), with its value ranging from 1 Byte up to 2048 Bytes. The increment option for watermark value is every 1 Byte unit.

The above mentioned 3 options of network card mode create a combinatorial problem that we chose to solve in [1] using GA. It is considered combinatorial problem because every distinct transmission has its unique optimal setting that will generate near-maximum throughput, the optimal setting itself will be taken from one of the 3 network card modes, which specially for passive wait and watermark mode have their own optimization of value with the ranges that have been previously mentioned.

There were 47 simulated network cards in [1] that represented 47 nodes/servers, which had gone through simultaneous multiple network cards optimization, according to their characteristics of received transmissions. Each of these network cards were optimized using GA.

The 36 implemented GA trials for every single network card consisted of 36 different settings. Every trial had different combination of population size and number of generations. The population sizes varied from 50 until 100 with 10 unit increment. And every single population size had varieties of number of generations starting from 50 up to 100 with also 10 unit increment. Additionally, all GA trials applied same mutation probability of 0.2 and crossover probability of 0.9.

The fluctuation condition itself was recorded if the number of individuals that achieved fitness value of 90 Mbps and above was less than 90% of the population size.

3 Experiment Results

This section presents the fluctuation analysis of all the 36 GA trials for each of 47 network cards. The analysis consists of the lowest, highest, and the average fluctuation point.

The ‘P’ symbol in the figure names represents population size, while the ‘G’ symbol represents generation size.

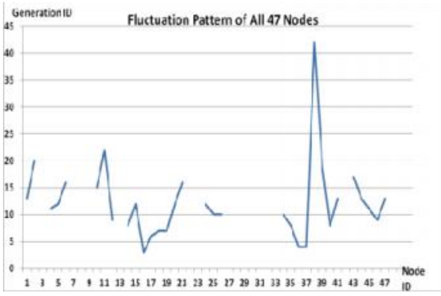

Fig. 1. Fluctuations in P = 50, G = 50

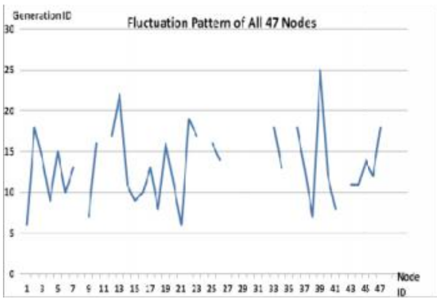

Fig. 2. Fluctuations in P = 50, G = 60

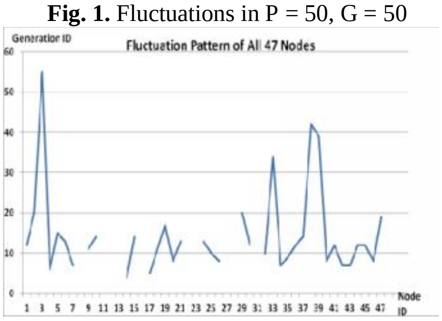

Fig. 3. Fluctuations in P = 50, G = 70

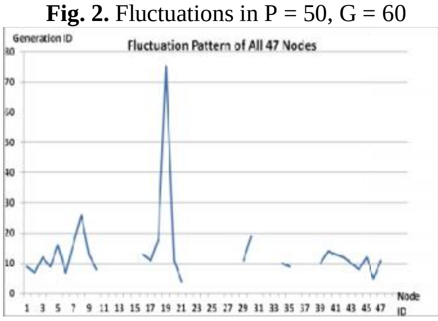

Fig. 4. Fluctuations in P = 50, G = 80

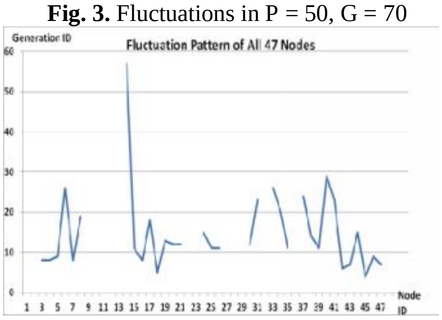

Fig. 5. Fluctuations in P = 50, G = 90

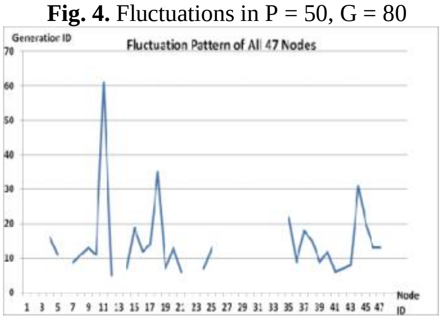

Fig. 6. Fluctuations in P = 50, G = 100

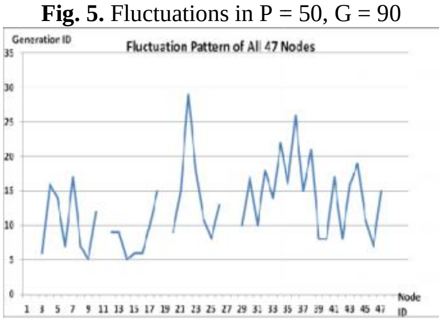

Fig. 7. Fluctuations in P = 60, G = 50

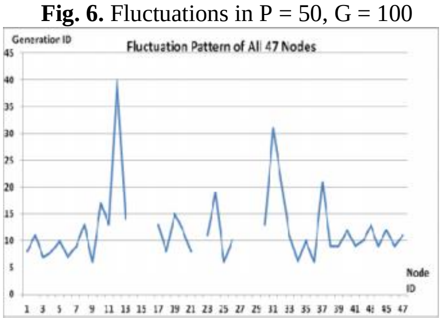

Fig. 8. Fluctuations in P = 60, G = 60

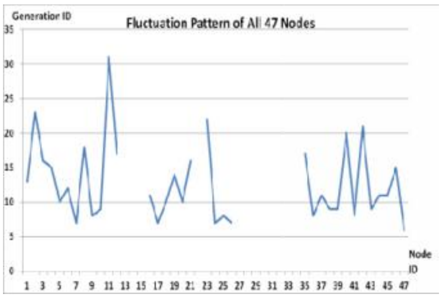

Fig. 9. Fluctuations in P = 60, G = 70

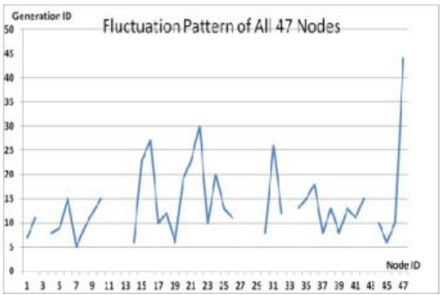

Fig. 10. Fluctuations in P = 60, G = 80

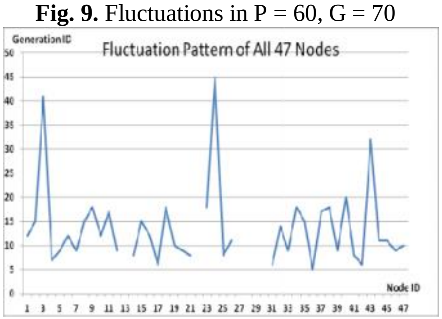

Fig. 11. Fluctuations in P = 60, G = 90

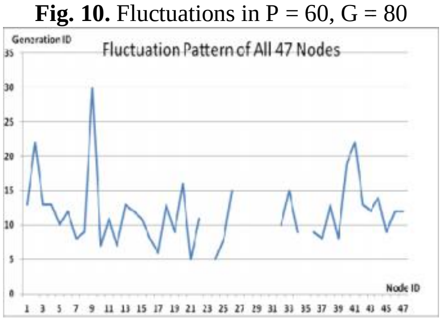

Fig. 12. Fluctuations in P = 60, G = 100

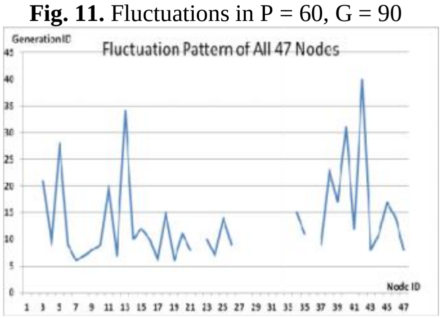

Fig. 13. Fluctuations in P = 70, G = 50

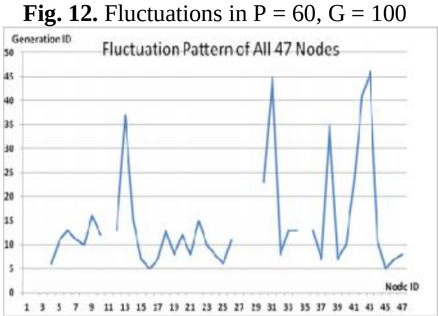

Fig. 14. Fluctuations in P = 70, G = 60

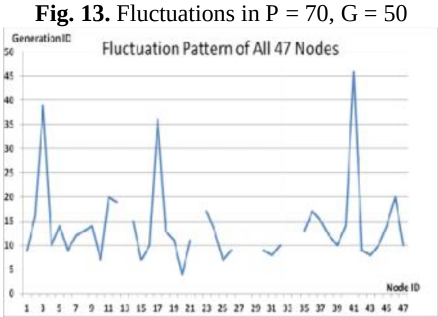

Fig. 15. Fluctuations in P = 70, G = 70

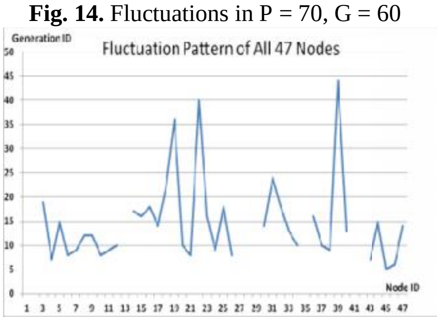

Fig. 16. Fluctuations in P = 70, G = 80

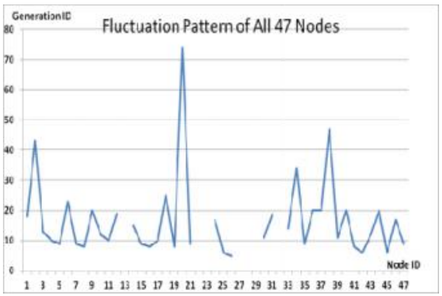

Fig. 17. Fluctuations in P = 70, G = 90

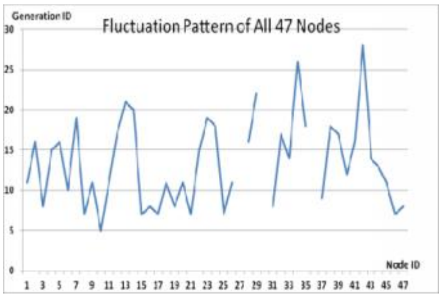

Fig. 18. Fluctuations in P = 70, G = 100

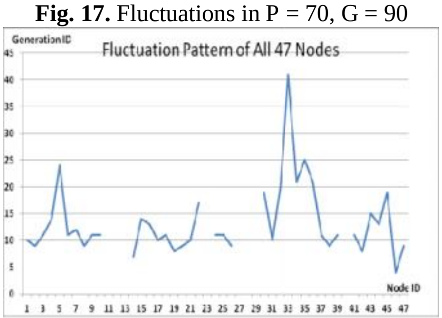

Fig. 19. Fluctuations in P = 80, G = 50

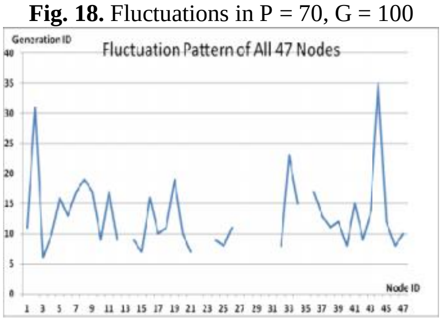

Fig. 20. Fluctuations in P = 80, G = 60

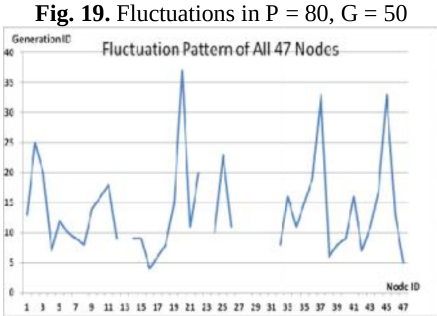

Fig. 21. Fluctuations in P = 80, G = 70

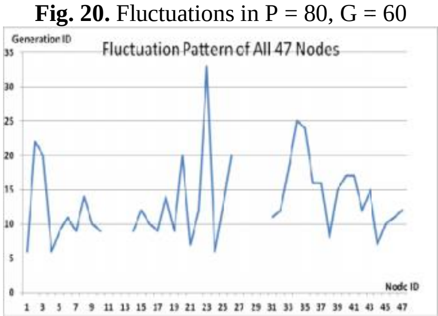

Fig. 22. Fluctuations in P = 80, G = 80

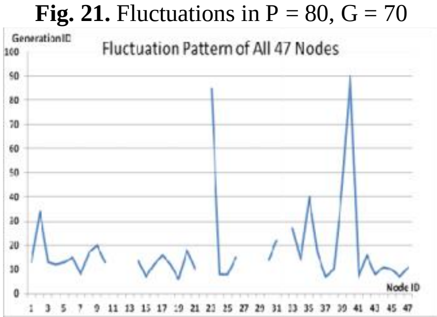

Fig. 23. Fluctuations in P = 80, G = 90

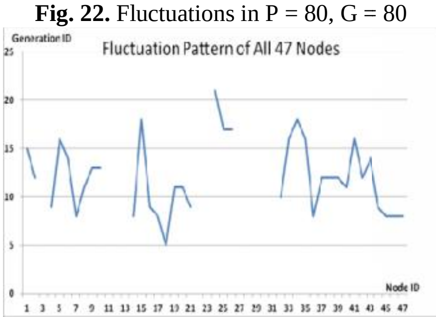

Fig. 24. Fluctuations in P = 80, G = 100

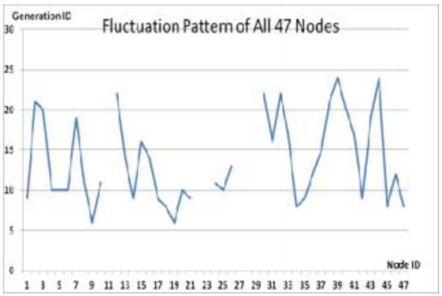

Fig. 25. Fluctuations in P = 90, G = 50

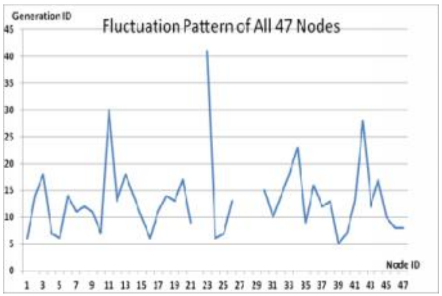

Fig. 26. Fluctuations in P = 90, G = 60

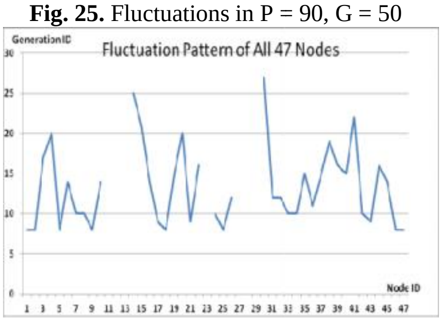

Fig. 27. Fluctuations in P = 90, G = 70

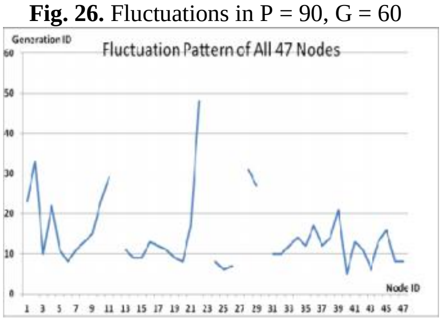

Fig. 28. Fluctuations in P = 90, G = 80

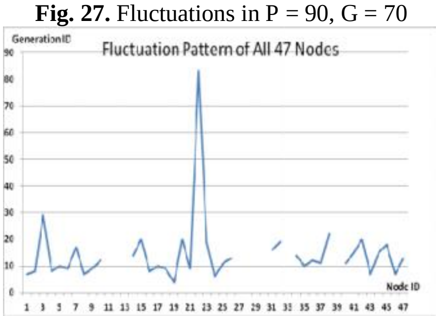

Fig. 29. Fluctuations in P = 90, G = 90

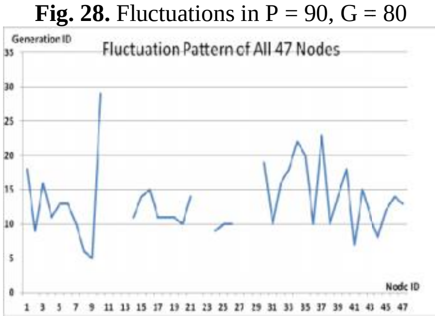

Fig. 30. Fluctuations in P = 90, G = 100

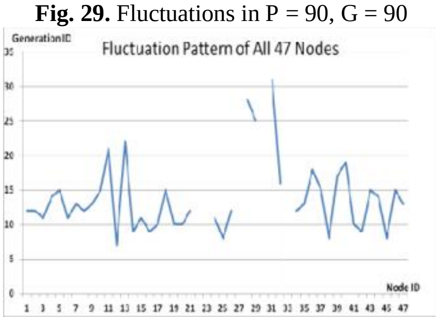

Fig. 31. Fluctuations in P = 100, G = 50

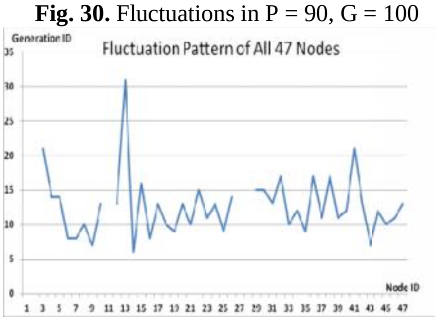

Fig. 32. Fluctuations in P = 100, G = 60

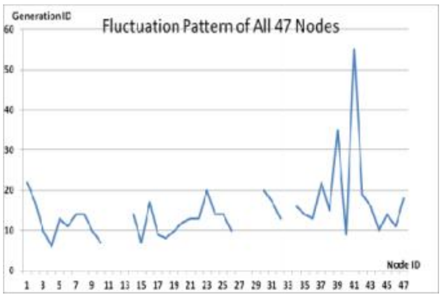

Fig. 33. Fluctuations in P = 100, G = 70

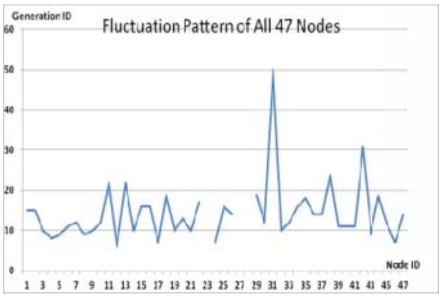

Fig. 34. Fluctuations in P = 100, G = 80

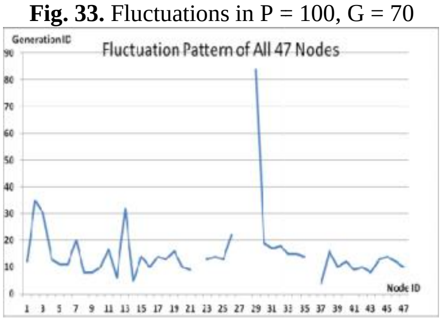

Fig. 35. Fluctuations in P = 100, G = 90

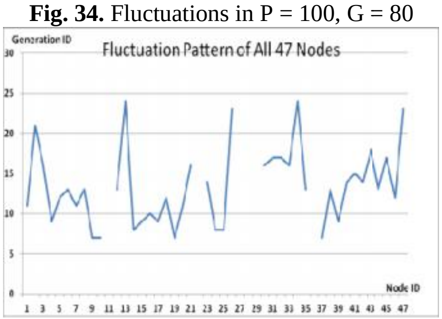

Fig. 36. Fluctuations in P = 100, G = 100

Table 1. GA Fluctuations Analysis Summary

Population & Generation Size	Lowest Fluctuation Point	Highest Fluctuation Point	Average Fluctuation Point
50, 50	4	42	12.37
50, 60	6	25	13.46
50, 70	4	55	14.39
50, 80	4	75	13.26
50, 90	4	57	15.05
50, 100	5	61	14.64
60, 50	5	29	12.67
60, 60	6	40	12.27
60, 70	6	31	13.37
60, 80	5	44	13.85
60, 90	5	45	13.67
60, 100	5	30	11.95
70, 50	6	40	13.85
70, 60	5	46	14.47
70, 70	4	46	14
70, 80	5	44	14.51
70, 90	5	75	16.57
70, 100	5	28	13.41
80, 50	4	41	13.32
80, 60	6	35	13.23
80, 70	4	37	14.12
80, 80	6	33	13.24
80, 90	6	90	18.35
80, 100	5	21	11.94
90, 50	6	24	13.68
90, 60	5	41	13.16
90, 70	8	27	13.22

90, 80	5	48	14.56
90, 90	4	83	14.46
90, 100	5	29	13.46
100, 50	7	31	13.83
100, 60	6	31	12.79
100, 70	6	55	14.90
100, 80	6	50	14.32
100, 90	4	84	15.49
100, 100	7	24	13.41

As we observe in Fig. 1 until Fig. 36, fluctuations did not always occur, which were identified by empty lines. However, in most cases it happened randomly without specific pattern. This could be due to mutation. These random fluctuations were also discovered by [8] in their convergence study. In our case, despite the frequent fluctuations, GA could generate best individuals for every node in all trials with near optimal fitness values starting from 99.9992 Mbps up to 100 Mbps as presented in [1].

Further analysis from Table 1 shows that the earliest fluctuations occurred at GA trials with small population, in this case 50, where 4 out of 6 cases have fluctuation points at 4th generation. This finding conforms to the nature of genetic algorithm itself, where small population tends to have less diversity that leads to smaller possibility of reaching near optimal solution.

Moreover, Table 1 also infers that there is no specific pattern of the high or late fluctuation points from all GA trials. Additionally, the average fluctuation points range from 12 up to 18th generation. It concludes that in general, fluctuation took place at the early stage of the genetic algorithm process.

4 Conclusions and Future Work

The findings of the behavior of GA in this paper have delivered the similarities between its implementation and inspiring nature, where small population is related to lesser diversity. However, its capability to find near-optimal solutions, in spite of random fluctuations displays its robustness and reliability in our case. This fact is expected to encourage other researchers to consider GA as their combinatorial optimization tool.

The related future work will be to study the fluctuation of GA in our another paper, which is about simulated packet clustering optimization in a data centre [9]. The study will compare the behavior of GA when an additional gene is added.

References

[1] Nurika, O., Hassan, M.F., Zakaria, N., Jung, L.T.: Genetic Algorithm Optimized Network in Cloud Data Centre. Advanced Science Letters. 22(10). pp. 2705-2709 (2015).

[2] Minsker, B.: Genetic algorithms. Hydroinformatics: Data Integrative Approaches in Computation, Analysis, and Modeling. P. Kumar, J. Alameda, P. Bajesy, M. Folk, and M. Markus, Eds. Florida, USA, CRC Press. pp. 439-456 (2005).

[3] Rogers, A., Prugel-Bennett, A.: Genetic Drift in Genetic Algorithm Selection Schemes. IEEE Transactions on Evolutionary Computation. vol. 3, no. 4 (1999).

[4] Gibbs, M.S., Dandy, G.C., Maier, H.R.: A genetic algorithm calibration method based on convergence due to genetic drift. Information Sciences 178. pp. 2857-2869 (2008).

[5] Dick, G., Whigham, P.: The behavior of Genetic Drift in a Spatially-Structured Evolutionary Algorithm. The 2005 IEEE Congress on Evolutionary Computation (Volume:2). pp. 1855-1860 (2005).
[6] Reed, P., Minsker, B., Goldberg, D.E.: Designing a competent simple genetic algorithm for search and optimization. Water Resources Research. vol. 36, no. 12. pp. 3757-3761 (2000).
[7] Nurika, O., Hassan, M.F., Zakaria, N., Jung, L.T.: Mathematical Models for Network Card Simulation and Their Empirical Validations. International Symposium on Mathematical Sciences & Computing Research (iSMSC2015) (2015).
[8] Louis, S.J., Rawlins, G.J.E.: Syntactic Analysis of Convergence in Genetic Algorithms. Foundations of Genetic Algorithms 1993 (FOGA 2). FOGA. Morgan Kauffman. pp. 141-151 (2014).
[9] Nurika, O., Hassan, M.F., Zakaria, N., Jung, L.T.: Genetic Algorithm Optimized Receiver-Oriented Packet Clustering in Multi-Buffer Network Card. Advanced Science Letters. 22(10). pp. 2914-2917 (2016).