Design and Performance Evaluation of a New Irregular Fault-Tolerant Multistage Interconnection Network

Rinkle Rani Aggarwal
Department of Computer Science & Engineering,
Thapar University, Patiala –147004 (India)

Abstract

Inter-connecting processors and linking them efficiently to the memory modules in a parallel computer is not an easy task. Hence, an interconnection network that provides the desired connectivity and performance at minimum cost is required. The design of a suitable interconnection network for inter-processor communication is one of the key issues of the system performance. The reliability of these networks and their ability to continue operating despite failures are major concerns in determining the overall system performance. In this paper a new irregular MIN named MASN (Modified Augmented Shuffle Exchange Network) has been proposed. The performance of MASN has been measured in terms of reliability and cost. It has been proved that the proposed network MASN provides much better fault-tolerance and reliability at lesser cost in comparison to ASEN-2.

Keywords: *Multistage Interconnection Network, Reliability, Augmented Shuffle Exchange Network, Fault-tolerance, MTTF.*

1. Introduction

Today is the era of parallel processing and building of multiprocessor system with hundred processors is feasible. Advances in LSI and VLSI technology are encouraging greater use of multiple-processor systems with processing elements to provide computational parallelism and memory modules to store the data required by the processing elements. Interconnection Networks (INs) play a major role in the performance of modern parallel computers. Many aspects of INs, such as implementation complexity, routing algorithms, performance evaluation, fault-tolerance, and reliability have been the subjects of research over the years. There are many factors that may affect the choice of appropriate interconnection network for the underlying parallel computing environment. Though crossbar is the ideal IN for shared memory multiprocessor, where N inputs can simultaneously get connected to N outputs, but the hardware cost grows astronomically. Multistage Interconnection Networks (MINs) are recognized as cost-effective means to provide programmable data paths between functional modules in multiprocessor systems [1]. These networks are usually implemented with simple modular switches, employing two-input two-output switching elements. Most of the MINs proposed in the literature have been constructed with 2x2 crossbar switches as basic elements, and have $n = \log_2 N$ switching stages with each stage consisting of N/2 elements, which makes the cost of this network as $O(N\log N)$, as compared to $O(N^2)$ for a crossbar [2]. The pattern of interconnection may be uniform or non-uniform, which classifies the MINs to be regular or irregular respectively. In the case of irregular networks, the path length varies from any input to any output, in contrast with regular networks, where it is the same. Fault-tolerance in an interconnection network is very important for its continuous operation over a relatively long period of time. Many networks have been designed and proposed to increase the fault-tolerance in the literature [3-7]. Permutation capability and other issues related to routing have also been extensively researched [8-10]. Various routing schemes have also been studied in-depth [11-14]. However, little attention has been paid to the computation of reliability of these networks. Reliability is measured in terms of Mean Time to Failure (MTTF), which is evaluated using simple series-parallel probabilistic combinations. This analysis is based upon the lower and upper bounds of the network reliability. This paper has been organized into five sections whose details are as follows.

Section 1 introduces the subject under study. Section 2 describes the structure and design of network. Section 3 focuses on the routing scheme. Section 4 concentrates on the reliability analysis of MASN network. Section 5 describes the cost effectiveness. Finally, the conclusion has been presented.

2. Design of Proposed Networks

Modified Augmented Shuffle Exchange Network (MASEN) is an irregular multistage interconnection network, derived from ASEN-2 network [15]. An $N \times N$ ($2^n \times 2^n$) network consists of m stages (where $m = \log_2 N/2$). The first and the last stage of the network contain equal number of switching elements that is 2^{n-1}, whereas the intermediate stages consist of less number of switching elements equal to 2^{n-2} each. The switches in the last stage are of size 2×2 whereas stages from 1 to m-1 are having switches of size 3×3. Thus, the total number of switches

are equal to $2^{n-2}(m+2)$ out of which 2^n number of switches are of size 2 × 2 and (m-2) × 2^{n-2} number of switches are of size 3 × 3 . There is one 4 × 1 multiplexer for each input link of a switch in first stage and one 1 × 2 demultiplexer for each output link of switch in the last stage. Hence, there exist 2N multiplexers and demultiplexers of size 4 × 1 and 1 × 2 respectively. MASEN of size 16×16 is shown in Figure 1.
Following structural changes have been made in ASEN-2:

1) Use of 4 × 1 MUX in place of 2 × 1 MUX.
2) Four switches removed from stage 1 (Intermediate Stage).
3) Change in loops and connections.

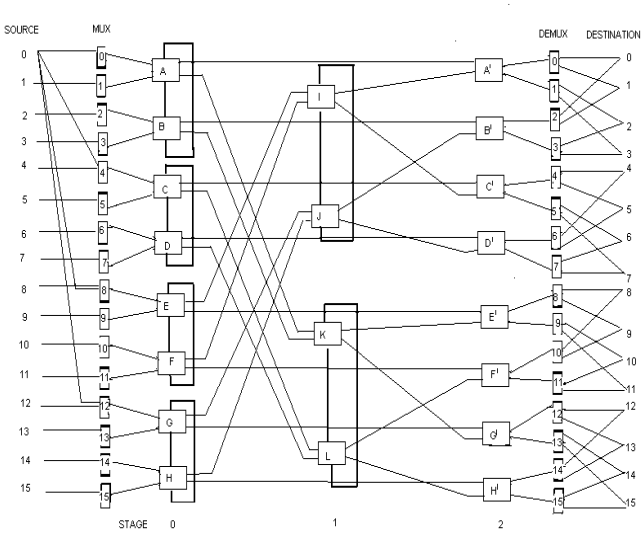

Fig. 1 MASN of size 16 x 16

3. Routing Scheme

3.1 Redundancy graph

A redundancy graph offers a convenient way to study the properties of a multi-path MIN, such as the number of faults tolerated or the type of rerouting possible. A redundancy graph also depicts all the available paths between a given source-destination pair within a MIN. It consists of two distinguished nodes source S and destination D. The rest of the nodes correspond to the switches that lie along the paths between S and D as shown in Figure 2.

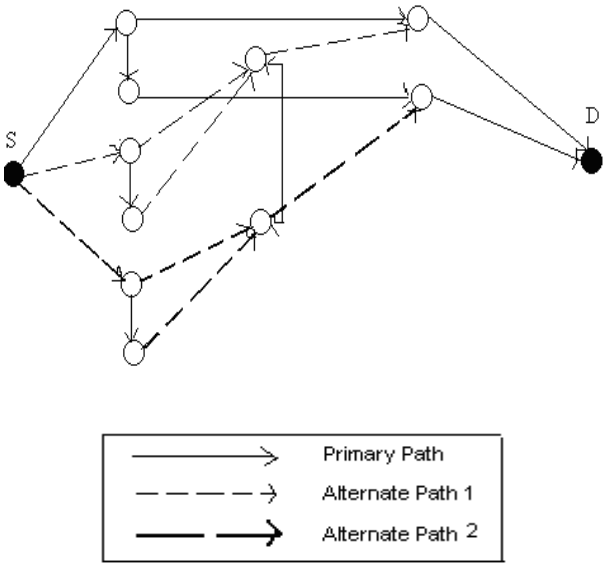

Fig. 2 Redundancy Graph for MASN Network

3.2 Routing Procedure

In MSAEN binary address of the destination is used as routing tag. For example let the source S and destination D are represented in binary code as:

$S=s_0, s_1 \ldots s_{n-2}, s_{n-1}$
$D=d_0, d_1 \ldots d_{n-2}, d_{n-1}$

Then at any stage, if d_i (where i=0 to n-1) is equal to 0 then follow the upper link otherwise follow the lower link.
A request from source S to a given destination D is routed through the MASN as:

i The source S selects one of the sub network G^i based on the most significant bit of the destination D ($i=d_0$).
ii Each source attempts an entry into the MASN via its primary path. If the primary path is faulty (i.e. either multiplexer or primary switch or both are faulty), then the request is routed to the alternate path.
iii To route a request through a network there are favourable path and less favourable paths. The path length algorithm is used to determine whether a request can be routed through the most favourable path or not. If most favourable path is not available or is busy then there is need of alternate path. If the alternate path is also not available, because it is busy or faulty, then drop the request.

iv A fault in the demultiplexer at the output of a switch in a stage (2m - 1) is regarded as a fault in that switch. From the demultiplexer, the request is routed to the upper or lower destinations according to the least significant bit of the tag (i.e. d_{n-1}).

3.3 Path length algorithm

If (S is even)
Then
If (S = D or D = S+1 or S-D = ± 8)
Then
Path length is minimum.
Else If (S is odd)
Then
If (S = D or D = S-1 or S-D = ± 8)
Then
Path length is minimum.
Else
The longest path is possible only.

Routing Tag →

If (path length is minimum)
Then
Routing tag = $0.D_1.D_{n-1}$
Else
Routing tag = $D_0.D_1.D_2.D_{n-1}$

3.4 Implementation

In order to find all the possible paths between a given source-destination pair an algorithm in 'C' has been designed. The algorithm has used following methodology.

a) The MINs have been realized as directed graph. All the switches, multiplexers and demultiplexers in the network are considered as nodes and are numbered from 0 to n-1, where n is the total number of nodes in the directed graph.
b) The topology of the various MINs has been stored as an adjacency matrix using sequential memory representation.
c) The Dijkastra's shortest path algorithm has been used to find the most favourable path i.e. the primary path in case multiple paths are available.
d) In the presence of fault, instead of reinitializing the request, the concept of dynamic re-routing has been used to find an alternate path from the current stage. If no such alternate path is available in the forward direction then backtracking to the previous stage has been done to explore any other path possibility.

Example: Let Source=0000 and Destination=0000 then Following paths are possible between source 0000 and destination 0100 as shown in Figure 3.

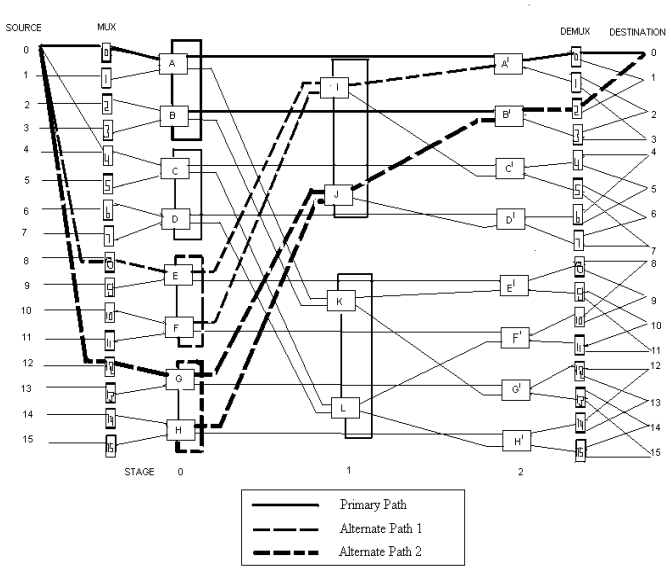

Fig. 3 Possible Paths between (0,4) in MASEN

Primary path :

0 →MUX (0) → A → A' → DEMUX (0) →0

Alternate Paths:
0 → MUX (0) → A → B → B' → DEMUX (2) →0
0 →MUX(8) → E → I → A'→ DEMUX(0) →0
0→ MUX (8) → E → F → I → A'→DEMUX (0) → 0
0→ MUX (12) → G →J → B' → DEMUX (2) →0
0→ MUX (12) → G → H → J→ B' → DEMUX (2) → 0

From Figure 3 it is clear that there exist six paths between a given source-destination pair in MASN network. Whereas, in case of ASEN-2 there exists only 2 paths. Therefore the proposed network MASN can entertain more number of requests even under faults in comparison to ASEN-2. Thus MASN is more fault-tolerant than ASEN-2.

4. Reliability Analysis

Reliability of MASN network is analyzed in terms of Mean time to Failure (MTTF) using simple series-parallel probabilistic models. This analysis is based upon the lower and upper bounds of the network reliability. The assumptions used in the analysis on the failure rates of the components are given below:

i Switch failure occurs independently in a network with a failure rate of λ for 2x2 crossbar switches (a reasonable estimate for λ is about 10^{-6} per hour).
ii Failure of the multiplexers and demultiplexers also occur independently with failure rates of λ_m and λ_d respectively.

iii Assuming that the hardware complexity of a component is directly proportional to the gate counts of it, the number of gates in a 2x2 crossbar switch is approximately equal to that in a 2x1 MUX or a 1x2 DEMUX. Thus $\lambda_m = \lambda_d = m\lambda/2$ for a mx1 MUX, where λ_m and λ_d are failure rates of MUX and DEMUX respectively.

iv The adaptive routing scheme considers a 2x2 switch in the last stage and its associated DEMUX as a series system, so consider these three elements as single component (SE2d), and assign failure rate of $\lambda_{2d} = 2\lambda$ to this group of elements. Also let λ_2 and λ_3 are the failure rates for the 2x2 and the 3x3 switches respectively then, $\lambda_2 = \lambda$ and $\lambda_3 = 2.25\lambda$ and $\lambda_{3m} = 4.25\lambda$.

v Irregular MINs are inherently multi-path and the MTTF needs to be calculated at all existing path-lengths separately based upon the series and parallel models of reliability.

4.1 Optimistic (Upper Bound) Analysis of MASN Network

In MASN each source is connected to two multiplexers and each SE has a conjugate pair. So it is assumed that the MASEN is operational as long as one of the two multiplexers attached to a source is operational and as long as both components in a conjugate pair are not faulty. The block diagram is shown in Figure 4.

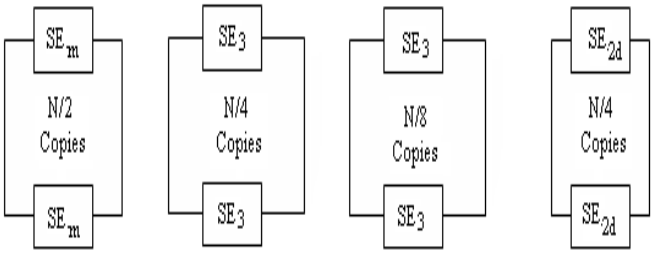

Fig. 4: Block diagram of Upper Bound MASEN

Reliability Equations are:

$$f1 = \left[1 - \left(1 - e^{-\lambda_m t}\right)^2\right]^{(N/2)}$$

$$f2 = \left[1 - \left(1 - e^{-\lambda_3 t}\right)^2\right]^{(N/4)}$$

$$f3 = \left[1 - \left(1 - e^{-\lambda_3 t}\right)^2\right]^{(N/8)}$$

$$f4 = \left[1 - \left(1 - e^{-\lambda_{2d} t}\right)^2\right]^{(N/4)}$$

$$R_{Optimistic} = f1 * f2 * f3 * f4$$

$$MTTF = \int_0^\infty R_{Optimistic}(t)\, dt$$

4.2 Pessimistic (Lower Bound) Analysis of MASEN Network

At the input side of the MASEN, the routing scheme does not consider the multiplexers to be an integral part of a 3 × 3 switch. For example, as long as at least one of the two multiplexers attached to a particular switch is operational, the switch can still be used for routing. Hence, if two multiplexers are grouped with each switch in the input side and consider them a series system (SE_{3m}), their aggregate failure rate will be $\lambda_{3m} = 4.25\lambda$. Finally these aggregated components and the switches in the intermediate stages can be arranged in pairs of conjugate loops. To obtain the pessimistic (lower) bound on the reliability of MASEN, assume that the network is failed whenever more than one conjugate loop has a faulty element or more than one conjugate switch in the last stage fails. The block diagram is shown in Figure 5.

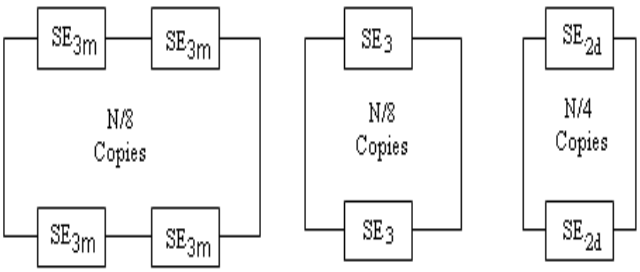

Fig. 5: Block diagram of Lower Bound MASEN

Reliability Equations are:

$$f1 = \left[1 - \left(1 - e^{-\lambda_{3m} t}\right)^2\right]^{(N/8)}$$

$$f2 = \left[1 - \left(1 - e^{-\lambda_3 t}\right)^2\right]^{\left(N/8\right)}$$

$$f3 = \left[1 - \left(1 - e^{-\lambda_{2d} t}\right)^2\right]^{\left(N/4\right)}$$

$$R_{Pessimistic} = f1 * f2 * f3$$

$$MTTF = \int_0^\infty R_{Pessimistic}(t)\,dt$$

The results of the MTTF Reliability equations have been shown in Table 1.

Table 1: MTTF of FT and IFT for different Network Size

Network Size (LogN)	ASEN-2		MASEN	
	Lower Bound	*Upper Bound*	*Lower Bound*	*Upper Bound*
4	5.07329	5.13012	5.99863	5.86775
5	4.84479	4.89034	5.78921	5.67442
6	4.62714	4.67522	5.58937	5.48681
7	4.44248	4.47502	5.39675	5.3033
8	4.25612	4.28454	5.2095	5.12275
9	4.07555	4.10075	5.02618	4.94431
10	3.89916	3.92184	4.84571	4.76736

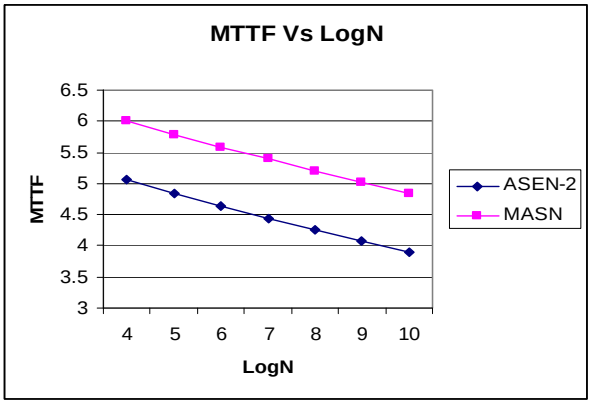

Fig. 6 MTTF (Lower Bound) comparison of ASEN-2 and MASN

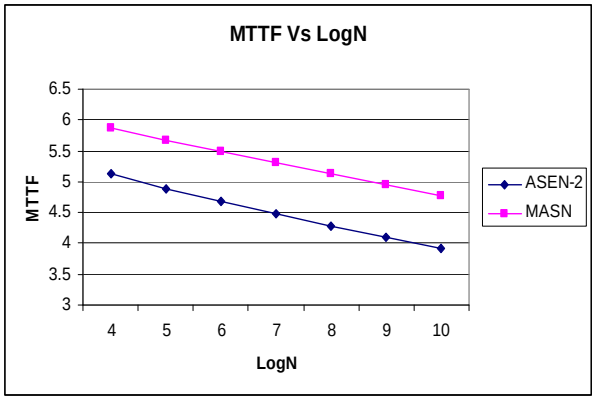

Fig. 7: MTTF (Upper Bound) comparison of ASEN-2 and MASN

Figures 6 and 7 depict that the proposed network MASN is more reliable than the existing ASEN-2 network for both the upper and lower bounds of reliability.

5. Cost Analysis

To estimate the cost of a network it has been assumed that the cost of a switch is proportional to the number of cross-points within a switch [16]. For example a 4x4 switch has 16 units of hardware cost whereas a 2x2 switch has 4 units. The cost functions for the proposed and existing MINs are given in the Table 2. From Figure 8 it is clear that MASN and ASEN-2 both are having comparable cost.

Table 2: Cost Functions for Networks

Network	Cost Function
ASEN	$3N(1.5 \log_2 N - 1)$
MASEN	$3N(1.5 \log_2 N) - 52$

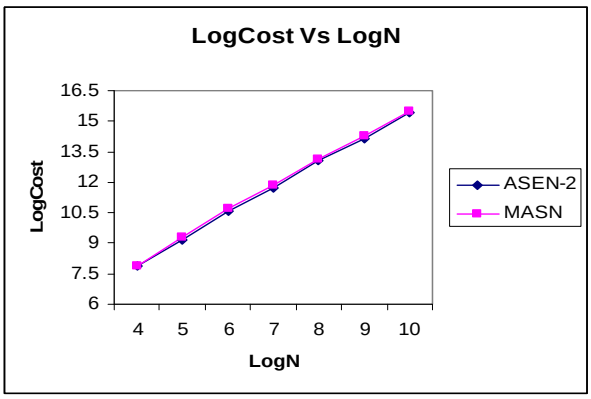

Fig. 8 Cost Comparison of ASEN-2 and MASN

6. Conclusion

Modified Augmented Shuffle Exchange Network (MASN) is designed using existing Augmented Shuffle Exchange Network (ASEN-2). It has comparatively lesser number of switches in the intermediate stage and thus reduced cost than ASEN-2. MASN is a dynamically re-routable network that provides multiple paths of varying lengths between a source-destination pair. It has been found that in MASN, there are six mutually exclusive distinct possible paths between any source-destination pair, whereas ASEN-2 has only two such paths. Thus the new network MASN can entertain larger number of requests even under faults. The upper and lower bound reliability analysis shows that MASN is more reliable than ASEN-2. Thus the new network MASN provides better fault-tolerance and reliability than the existing ASEN-2 without any additional cost.

References

[1] L.N. Bhuyan, Y. Qing and D.P. Agrawal, "Performance of Multiprocessor Interconnection Networks", Computer, Vol. 22, No. 2, 1989, pp. 25-37.

[2] A. Gupta and P.K. Bansal, "Evaluation of Fault-tolerant Multistage Interconnection Networks", CSI-Communication, 2008, pp. 1-16.

[3] P. K. Bansal, K. Singh, R.C. Joshi and G.P. Siroha, "Fault-tolerant Augmented Baseline Multistage Interconnection Network", IEEE International Conference TENCON-91, 1991, pp. 200-204.

[4] G. Hou and Y. Yang, "Super Recursive Baselines: A Family of New Interconnection Networks with High Performance/Cost Ratios", IEEE International Symposium on Parallel Architectures, Algorithms and Networks ISPAN- 2000, 2000, pp. 260-265.

[5] Nitin, "On Analytic Bounds of Regular and Irregular Fault-tolerant Multistage Interconnection Networks", International Conference PDPTA-06, 2006, pp. 221-226.

[6] J. Sengupta and P.K. Bansal, "High speed Dynamic Fault-Tolerance", IEEE Region 10 International Conference on Electrical and Electronic Technology, 2001, pp. 669-675.

[7] S. Sharma, K.S. Kalhon, P.K. Bansal and K. Singh, "Improved Irregular Augmented Shuffle Exchange Multistage Interconnection Network", International Journal of Computer Science and Security, Vol. 2, No. 3, 2008, pp. 28-33.

[8] J. Sengupta and P.K. Bansal**, "**Performance of Regular and Irregular Dynamic MINs", IEEE International Conference TENCON-99, 1999, pp. 427-430.

[9] S. Sharma, K.S. Kahlon and P.K. Bansal, "On a class of Multistage Interconnection Networks in Parallel Processing", International Journal of Computer Science and Network Security, Vol. 8, No. 5, pp. 287-291.

[10] A. Subramanyam, E.V. Prasad and R. Nadamuni, "Permutation Capability and Connectivity of Enhanced Multistage Interconnection Network (E-MIN)", IEEE International Conference on Advanced Computing and Communications, 2006, pp. 8-11.

[11] S. Seok, B. You, S. Youm, K. Kim and C. Kang, "A Heuristic Multi-path routing scheme for online traffic in MPLS Networks", International Journal of Computer Systems Science and Engineering, Vol. 25 No. 1, 2010.

[12] Y. Tang, Y. Zhang and H. Chen, "A parallel shortest path algorithm based on graph-partitioning and iterative correcting", International Journal of Computer Systems Science and Engineering, Vol. 24, No. 5,2009.

[13] C. Zhen, L. Zengji, Q. Zhiliang, C. Peng and T. Xiaoming, "Balance Routing Traffic in Generalized Shuffle-Exchange Network", Journal of Electronics, Vol. 22, No. 4, 2005, pp. 345-350.

[14] P. K. Bansal, K. Singh and R.C. Joshi, "Routing and path length algorithm for a cost-effective Four-Tree Multistage Interconnection Network", International Journal of Electronics, Vol. 73, No.1, 1992, pp. 107-115. DOI: 10.1080/00207219208925650

[15] J. Sengupta, P.K. Bansal and A. Gupta, "Permutation and Reliability Measures of Regular and Irregular MINs", IEEE International Conference TENCON-2000, 2000, pp. 531-536.

[16] H. Aggarwal and P.K. Bansal, "Routing and Path Length Algorithm for Cost-effective Modified Four Tree Network", IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering, 2002, pp. 293-297.

Dr. Rinkle Aggarwal is working as Assistant Professor in Computer Science and Engineering Department, Thapar University, Patiala since 2000. She has done her Post graduation from BITS, Pilani and Ph.D. from Punjabi University, Patiala in the area of Parallel Computing. She has more than 14 years of teaching experience. She has supervised 20 M.Tech. Dissertations and contributed 33 articles in Conferences and 25 papers in Research Journals. Her areas of interest are Parallel Computing and Algorithms. She is member of professional bodies: ISTE, IAENG, IACSIT & WASET and ICGST.