Indian Journal of Science and Technology, Vol 9(31), DOI: 10.17485/ijst/2016/v9i31/97805, August 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

An Optimum Value of Dynamic Communication Performance for Midimew Connected Mesh Network

Ala Ahmed Yahya Hag[1*], M.M. Hafizur Rahman[1], Rizal Mohd Nor[1] and Tengku Mohd Tengku Sembok[2]

[1]Department of Computer Science, KICT, IIUM, Jalan Gombak, 53100, Kuala Lumpur, Malaysia; ala_hag500@yahoo.com, hafizur@iium.edu.my, rizalmohdnor@iium.edu.my
[2]Cyber Security Center, National Defense University Malaysia, Kuala Lumpur 57000, Malaysia; tmtsembok@gmail.com

Abstract

A Midimew connected Mesh Network (MMN) is a proposed hierarchical interconnection network in which numerous basic modules are interconnected in a hierarchical fashion to construct a massively parallel computer. Here, the basic module is the first level of the network which is connected by mesh network while higher level network is connected by a midimew network. In this study, we have evaluated a dynamic communication performance of a variety of MMN using the TOPAZ simulator using a deadlock free routing algorithm with uniform traffic patterns, whereby the dimension order routing and virtual cut-through flow control are used. It is found that the saturation throughput of the Virtual MMN (VMMN) is higher than that of Horizontal MMN (HMMN), Symmetric MMN (SMMN), and Tori connected mESH network (TESH).

Keywords: Dynamic Communication Performance, HMMN, Massively Parallel Computer, SMMN, VMMN

1. Introduction

Advancement in hardware technology, especially in *Very-Large-Scale Integration* (VLSI) circuit and network-on-chip design makes it possible to build parallel computers possible to solve the existing and forthcoming demand of computational power. As the increasing demand will never stop, we need extraordinary computing power to solve the forthcoming computationally challenging problems which is also known as grand challenge problems in the science and technology fields[1]. These include weather forecasting, health care improvement, medicines development, and disaster mitigation, and so on. Therefore, we need a special computer system which will yield petaflops or exaflops level of computing performance. This type of computer system is known as Massively Parallel Computer (MPC) systems. A MPC system can solve these computational intensive grand challenge problems in rational time. To achieve this massive performance, a MPC system consists of millions of nodes[2].

To design such a MPC system, the first vital step is to decide the interconnection network topology of the system, because it affects the whole performance of a network system[2,5]. Currently the conventional topologies are widely used in commercial parallel computers. In very near future, these conventional topologies with very large diameter will be absolutely implausible to build a MPC system consisting of millions of nodes[6,8]. However, it is believed that the hierarchical interconnection network is a conceivable way in building future generation MPC system[2]. Since several network topologies can be integrated together in one network for better cost effective design. This is why, the use of *Hull Identification Number* (HIN) to build future MPC systems will be a great solution.

In the literature, many theoretically attractive hypercube based hierarchical interconnection network is found, however, practically none of them is suitable because these networks have a very high number of links[3,4]. Then researchers are trying to find the alternative

*Author for correspondence

hierarchical interconnection network using k-ary-n-cube network. Though it has some potential advantage but still it did not draw the attention from the industry community. To find an optimal hierarchical interconnection network, the researcher are trying to find a suitable hierarchical network which will yield better performance in as maximum aspect as possible with potential reduction of implementation cost and improvement of scalability. It is to be noted that the interconnection network cost is 25% to 30% of the total cost of a MPC system[7].

A Tori connected mESH named as TESH is an example of hierarchical interconnection network using k-ary n-cube network[3]. It was proposed for future generation MPC systems. This network is made up of modular implementation. i.e., a 2D-mesh network refers as a level-1 or basic module whereas a multiple basic modules of 2D-mesh are interconnected using 2D-torus network. The free ports along with the links in the periphery of the basic module are used for the interconnection of higher level TESH network. However, TESH network results low saturation throughput due to lack of connectivity in the higher level networks. Midimew network yield low latency and high throughput because of its diagonal wrap-around connection. Midimew network yields low diameter and average distance among all networks whose node degree is four.

Considering these results of midimew network, we have replaced the toroidal connection of higher level TESH network by the midimew network. Likewise the TESH network, the free ports and the associated additional links in the basic module used to interconnect higher level network using midimew connection. The Midimew connected Mesh Network (MMN) consisting of various basic modules (BM) wherein the BM is a 2D-mesh network. And these BMs are interconnected together a hierarchical fashion using minimal distance mesh with diagonal wrap-around links (also known as midimew network) to form higher level MMN. The main focus of this paper is to evaluate the Dynamic Communication Performance (DCP) of a MMN and its variation Horizontal MMN (HMMN) and Vertical MMN (VMMN)[5,7,8]. That is, the main objective of this paper is to find the variation of MMN which yields the optimum DCP, i.e., lowest latency and highest throughput among all variation SMMN, HMMN, and VMMN.

The rest organization of the paper is as follows. The basic structure of the MMN and its derivatives are discussed in Section 2. Node addressing and routing of message in the MMN are explained in Section 3. The DCP evaluation of MMN and its derivatives are explained in Section 4. Finally, Section 5 is the conclusion of this paper.

2. Structure of the MMN

A Midimew connected Mesh Network (MMN)[7,9] is an hierarchical interconnection network having numerous BMs whereby the BM is a mesh network of 2D type and the higher-level of hierarchy are formed by connecting the BMs using midimew network. The size of BM of mesh network type 2D consist of $(2^m \times 2^m) = 2^{2m}$ nodes having 2^m rows and 2^m columns in grid form an it consider as level 1 network. Here m is any value of positive integer. Yet, the superior choice is m = 2 for better granularity. Thus, seeing m as 2, a (4×4)-size BM is illustrated in Figure 1(a). In each outlines of the BM, there are 2^{m+2} numbers of free-port which are used in interconnection for higher level.

The free-port are usually used in interconnection in higher level as explained and illustrated in Figure 1(a). Higher-level network is constructed by recursively connecting (2^{2m}) of direct lower-level networks as sub-level network. Since, m is 2, a Level 2 could be constructed by linking $2^{2\times2} = 16$ of Level 1 (BMs), a Level 3 MMN could be shaped by interrelating 16 of Level 2, and so on.

Figure 1(b) shows a clear view of this phenomenon. To avoid such confusion, the wraparound links of high level connections of the BMs are not illustrated. Every BM has $2^2 \times (2^q) = 2^{q+2}$ of its free-link, where is half of that (2^{q+1}) free-link used in vertical connections and the second half of the (2^{q+1}) free-link used in horizontal connections. A new symbol (q) represent the inter-level connectivity, where $0 \le q \le m$. when q = 0 means the smallest inter-level connectivity, while q = m is the largest inter-level connectivity. The outlet nodes of Basic modules are linked to build a corresponding network in higher-level using the corresponding allocated free-link. These free links assignments for incoming and outgoing links are the key difference for designing the variety of MMN such as Symmetric MMN, Horizontal MMN and Vertical MMN. Each of these SMMN, HMMN and VMMN use different arrangement of the free ports assignment to connect BMs forming the higher-levels of the network (5, 8). Figure 1(a) demonstrated a (4′4) basic module considering m = 2 resulting of $2^{2+2} = 16$ free ports. As q = 0, $4(2^0) = 4$ quantity of free-port are used for both vertical and horizontal connections to shape higher level interconnection

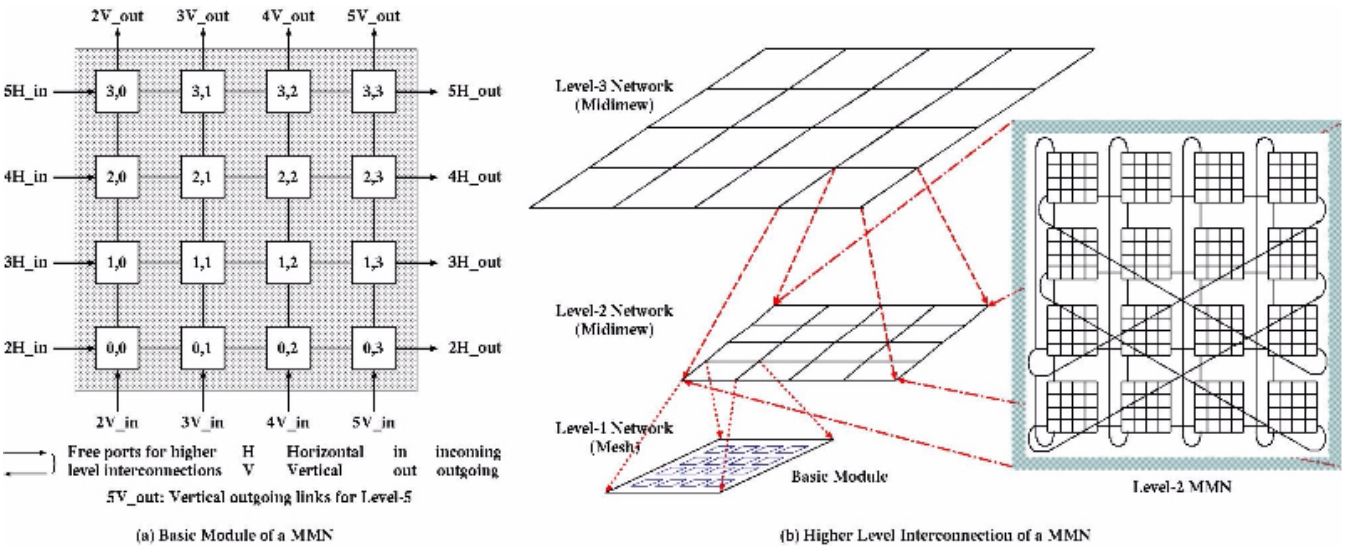

Figure 1: Interconnection of a SMMN (a) Basic module (b) Higher level network

in each level. The 4 free ports in each direction are split into halves, 2 free ports are used for incoming connection and another 2 free ports used for outgoing connections. One incoming and one outgoing links are tied together to form a b-direction links to connect two adjacent basic modules to form higher level networks. Both vertical and horizontal directions used incoming and outgoing links as depicted at the periphery of the basic models to be connected in a form of midimew connections as shown in Figure 1(b) Level-2.

To build higher level network of MMN by considering number level of hierarchy L along with m and q, MMN is symbolized as MMN (m, L, q). It is created using 2^{2m} BMs using L levels of hierarchy and q interlevel connectivity. In this paper, we have assigned m = 2, the best granularity. As a result, we focus on MMN (2,L, q) networks. With m= 2, the maximum possible level of hierarchy, L_{max}, of a MMN can be build using (2^{2m}) BM is $L_{max} = 2^{m-q} + 1$. Considering, q = 0 and m = 2, $L_{max} = 2^{2-0} + 1 = 5$. Level-5 is the uppermost likely level that could be shaped in MMN through the (4×4) BM interconnection and its connected free-port and links. The full number of nodes of a Level L MMN is $N = 2^{2mL}$. With L=5, q=0 and m=2, MMN (2, 5, 0), $N = 2^{2m(2^{m-q}+1)}$ where N is the largest amount of nodes in the network. MMN with these parameters can consist huge number of nodes, grow up to millions of nodes that are interconnected in hierarchy fashion to form a massively parallel computer systems.

3. Node Addressing and Message Routing in MMN

3.1 Nodes Addressing

Each node of MMN network is represented by 2 digits, as shown in equation.1, these pairs of digits are assigned for the row index and column index. Usually, in level L of MMN, the addressing nodes can be signified by the equations below.

$$\begin{aligned} A &= A^{l} A^{l-1} A^{l-2} \ldots A^{2} A^{1} \\ &= a_{n-1} a_{n-2} a_{n-3} \ldots a_{2} a_{1} a_{0} \\ &= a_{2l-1}\ a_{2l-2} a_{2l-3} a_{2l-4} \ldots a_{3} a_{2} a_{1} a_{0} \\ &= (a_{2l-1} a_{2l-2} \text{ } a_{2l-3} a_{2l-4}) \ldots\ldots (a_{3} a_{2} \text{ } a_{1} a_{0}) \end{aligned} \quad (1)$$

The maximum number of digits in MMN network is n=2L, L represent a level number and A^L is the level address. A (a_{2L-1} a_{2L-2}) is the position of (L – 1) Level for Level-L network. Grouping the pair of digits to be executed from group one for Level-1, i.e., the BM, together number L for the level L-th. Explicitly, l-th group (a_{2l-1} a_{2l-2}) shows the position of a (l – 1) Level sub-network inside the l-th group where the node belongs; 1 ≤ l ≤ L. In a network of two-level, the address presented as A = (a_4 a_3) (a_1 a_0). These pair of digits (a_4 a_3) recognizes the basic module where the node belongs, and the pair of digits (a_1 a_0) identifies the exact node within that basic module.

Assigning inter-node ports for the higher-level interconnection network was accomplished cautiously in order to minimize the higher-level traffic through the BM. The

n^1 represents node address involves in BM_1 is written as $n^1 = (a^1_{2L-1}\, a^1_{2L-2}$ .... $a^1_3 a^1_2\, a^1_1 a^1_0)$. The node n_2 address included in BM2 is signified as $n^2 = (a^2_{2L-1}\, a^2_{2L-2}$ ..... $a^2_3 a^2_2\, a^2_1 a^2_0)$. The link between node n^1 in BM_1 and n^2 in BM_2 are interconnected only if it satisfies the condition below.

$$\exists i\{a_i^1 = (a_i^2 \pm 1) \bmod 2^m \\ \wedge \forall j\{j \neq i \rightarrow a_j^1 = a_j^2)\} where \cdots i, j \geq 2 \quad (2)$$

3.2 Routing Algorithm of MMN

The data communications between each node in the network through interchange messages depends completely in the routing algorithm, which determines the path for a message to go through channel from source node to destination node. Both, simplicity of router design and the effectiveness of message routing are vital to the performance and adopting interconnection network for immensely parallel computer systems.

Midimew connected Mesh Network is a hierarchical interconnected network, where it's routing of messages will go in different stages. In general, messages routing in MMN is performed using top-bottom approach like TESH network[4]. First, it is performed at the top level network; at that point, when the packet arrive its top-level sub-destination, routing lasts inside the sub-networks to the following lower-level sub-destination. This scenario repeated until the packet delivered to its ending destination. Once a packet is created at a source-node, the source-node looks for its final destination. If the destination of a packet is local (inside the current BM) then the routing will be carried out inside the BM only. If the address is external (different BM), the packet will be sent to the outlet node that links the BM to the level where the routing is supposed to continue. We have bear in mind the dimension order routing algorithm for the MMN due to easiness and inexpensive hardware requirements.

For every level, routing performs in two directions. First routing is done in a vertical direction (first in the y-direction). The routing in y-direction, endwise node is employed then the routing performed diagonally otherwise routing goes vertically. When the packet arrives at the correct row, then the routing is implemented in the horizontal direction (routing in the x-direction) like basic routing in torus. For dimension-order routing, the path of routing is determined by the addresses of source and destination nodes and that is sufficient to determine the pathway that a packet follows. MMN routing is entirely characterized by the addresses nodes of both source and destination.

4. Dynamic Communication Performance of MMN

In parallel computing systems, a problem is divided into different parts; each part is executed by each node of a parallel computer after that the result of each node is merged together to have the complete solution of the given problem; meaning that the coordination and cooperation among the nodes is very essential in a massively parallel computer system. This phenomenon is known as dynamic communication. Therefore, the dynamic communication performance is highly depending on two factors in a massively parallel computer system such as the communication performance of the underlying interconnection network as well as individual node's performance. Any matter within the interconnection network performance or individual nodes' performance can firmly affect the speed of the entire MPC system. As a result, the victory of building a high performance massively parallel computer system is greatly depending on the efficiency of communication of the interconnection networks.

4.1 Performance Metrics

The DCP is characterized by two parameters, viz. latency measured as average transfer time (simulator clock cycles) and throughput measured as number of flits per clock cycle per node. A MPC system is to be considered as good one, then low message latency and high network throughput that must be achieved.

4.2 Simulation Environment

To evaluate the DCP of an MMN and its derivative SMMN, HMMN and VMMN as well as its counter rivals TESH network, we have used an open source simulator named TOPAZ. Message routing is performed using the simple dimension order routing, which allocate the path for a packet to be transferred from source to destination node. For switching a message to packets after that packet to flits, virtual cut-through switching and flow control mechanism is used. To avoid any potential deadlock in the network during message routing, four virtual channels per physical link is used. For synthetic traffic, the uniform traffic pattern has been used; which is the most

efficient and widely used pattern whereby the source-destination pairs are randomly selected. That is, in this simulation, sending messages between every node in the network is done with equal probability. In this paper, we have considered Level-2 network consisting of 256 nodes is used for the DCP evaluation. Finally, we have considered SMMN, HMMN, VMMN and TESH network for the evaluation of DCP.

Each message is split into small packets and each packet is split into small unit called flits, the size of a packet is 5 flits, 2 flits for header and tail flits and 3 flits for data. Here, each flit is two byte length. We have simulated the networks for 20,000 cycles. For each physical link, four virtual channels are simulated and the arbitration of virtual channels is carried out by the round-robin algorithm. For the period of the evaluation of dynamic performance, herds of messages are transmitted over the network to compete for the output channels. The DCP is measured by means of changing the probability of packet generation. Therefore, for each interval of packet generation probability, we have recorded the latency as the average transfer time in simulation clock cycles and the network throughput as flits per cycle per node. Finally, we have plotted the average transfer time in the horizontal axis and throughput in the vertical axis.

4.3 Evaluation of Dynamic Communication Performance (DCP)

The evaluation results of the DCP of TESH, SMMN, HMMN, and VMMN have been evaluated according to the simulation environment as mentioned before and plotted the result in the Figure 2. This figure displayed the message latency (average transfer-time) in y-axis as a function of network throughput in x-axis. Every specific network is demonstrated in a specific curve. It can be comprehended that the latency of the MMN and its derivatives are marginally higher compare to the TESH network. While the injections of packets are increases in the network, the throughput and as well as latency of that network are also increased. In course of time, the situation of overcrowding packets is created in the network. Due to overcrowding of data in the network, the router buffer, i.e., the physical links and virtual channels become congested. The network becomes saturated and message latency increases drastically in the intervening time have no effect on network throughput. The network becomes saturated, whereby the network latency is exponentially increased but the throughput is not increased any more. The saturation point designates the maximum throughput in which a network reaches.

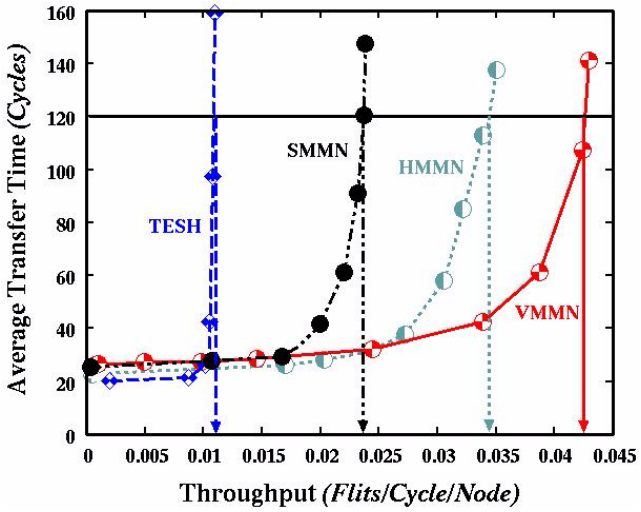

Figure 2: Dynamic communication performance of TESH, SMMN, HMMN, and VMMN using dimension-order routing with uniform traffic pattern and virtual cut-through flow control: 256 nodes, 4 VCs, 5 flits.

Figure 2 depicts clear representation of the evaluation of the DCP of TESH, SMMN, HMMN and VMMN. As depicted in Figure 2, we have seen that at 120 cycles all the networks are saturated. As shown that TESH network yields lowest DCP and it is about 0.01 *Flits/Cycle/Node*. All the variations of MMN (SMMN, HMMN, VMMN) yields high throughput during saturation at 120 cycles. However, the highest throughput yielded by VMMN and is about 0.042 *Flits/Cycle/Node*. That is, VMMN shows an excellent DCP for predictable future massively parallel computers. Meanwhile SMMN and HMMN present results in throughput compare to TESH. Thus, a VMMN is an excellent choice of hierarchical interconnection network for next generation of future massively parallel computer.

5. Conclusion

The DCP of a MMN and its two derivative HMMN and VMMN along with the TESH network is evaluated. We have found that MMN and its two variation yields low zero load latency and high throughput, i.e., better performance, especially in terms of throughput, than that of TESH network. The difference is zero load latency is trivial; however, the saturation throughput of all MMN

(VMM, HSMMN, and MMN) is superior to that of TESH network. Also, we found that among all the variation of MMN, VMMN yields an optimum saturation throughput over all the networks considered in this paper. We aim to future study and further explore the following: 1. reliability and fault tolerant performance of MMN, and 2. substitution of the long diagonal electronic link by the optical one, i.e., to explore the network structure and static and dynamic performance evaluation of op to-electronic MMN.

6. Acknowledgment

This research supported under grant FRGS13-065-0306, MOE, Malaysia. The authors are grateful for reviewers for their constructive suggestions' that assisted to enhance the quality of the research.

7. References

1. Youyao L, Jungang H, Huimin D. A Hypercube-based Scalable Interconnection Network for Massively Parallel Computing. Journal of Computers. 2008; 3:58-65.
2. Rahman MMH, Inoguchi Y, Sato Y, Horiguchi S. TTN: A High Performance Hierarchical Interconnection Network for Massively Parallel Computers. *IEICE* Transactions on Information and Systems. 2009; E92-D(5):1062-78.
3. Rahman MMH, Inoguchi Y, Sato Y, Miura Y, Horiguchi S. Dynamic Communication Performance of the TESH Network under Nonuniform Traffic Patterns. *Journal* of Networks. 2009; 4(10):941-51.
4. Jain VK, Horiguchi S. VLSI Considerations for TESH: A New Hierarchical Interconnection Network for 3-D Integration. *IEEE* Transactions on Parallel and Distributed Systems. 1998; 6(3):7.
5. Hag AAY, Rahman MM. Hafizur, Nor RM, Sembok TMT. On Uniform Traffic Pattern of Symmetric Midimew Connected Mesh Network. Procedia Computer Science Big Data. Cloud and Computing Challenges. 2015; 50:476–81.
6. Sibai F. A Two-Dimensional Low-Diameter Scalable On-Chip Network for Interconnecting Thousands of Cores. IEEE Transactions on Parallel and Distributed Systems. 2012; 23(2):193-1.
7. Hag AAY, Rahman MM. Hafizur, Nor RM, Sembok TMT, Miura Yc, Inoguchi Yd. Uniform Traffic Patterns using Virtual Cut-Through Flow Control on VMMN. Procedia Computer Science International Conference on Computer Science and Computational Intelligence (ICCSCI 2015). 2015; 59:400-9.
8. Hag AAY, Rahman MM. Hafizur, Nor RM, Sembok TMT. Dynamic Communication Performance of a Horizontal Midimew Connected Mesh Network *IJACT*. 2016; 8(1):31-40.
9. Rahman MMH, Shah A, Fukushi M, Inoguchi Y. Hierarchical Tori Connected Mesh Network. In: al BMe, editor. *ICCSA 2013*, Part V, LNCS 7975. Berlin Heidelberg Springer-Verlag. 2013; 197–210.