IJBPAS, July, 2015, 4(7): 4325-4335

ISSN: 2277–4998

International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)

'A Bridge Between Laboratory and Reader'

www.ijbpas.com

A NEW ARBITRATION METHOD TO IMPROVE ROUTING EFFICIENCY IN NETWORK-ON-CHIP

EBRAHIM BEHROUZIAN NEJAD[1*], HAMID MEHDI[2], AMIN EINIPOUR[3]

1: Department of Computer, Andimeshk Branch, Islamic Azad University,Andimeshk, Iran

2: Department of Computer, Andimeshk Branch, Islamic Azad University,Andimeshk, Iran

3: Department of Computer, Andimeshk Branch, Islamic Azad University,Andimeshk, Iran

***Corresponding Author: E Mail: Behrouzian.e@gmail.com**

ABSTRACT

Network-on-Chip (NoC) has been proposed as a solution to provide better modularity, scalability, reliability and higher bandwidth compared to bus-based communication infrastructures. The performance of Network-on-Chip largely depends on the underlying routing techniques. A routing technique has two constituencies: output selection and input selection.this paper focuses on the improvement of input selection. Two traditional input selections have been used in NoC, First-Come-First-Served (FCFS) input selection and Round-Robin input selection. Also, recently a contention-aware input selection (CAIS) technique has been presented for NOC, But there is some problem and defection in this technique. In this paper we improve the problems and defections of contention-aware input selection (CAIS) technique to develop a simple yet effective input selection technique named ICAIS. The simulation results with different traffic patterns show that ICAIS can achieves better performance than the FCFS and CAIS input selections, when combined with either deterministic or adaptive output selection.

Keywords: Network on chip, Routing Algorithm, Input selection

INTRODUCTION

In the past decades, System on Board (SOB) has been the dominant methodology for designing complex digital systems. As the complexity of applications and their required algorithms have grown so rapidly, SOB has been replaced by System on Chip (SOC)

methodology [1]. SOC consist of a number of pre-designed Intellectual property (IP) assembled together using electrical bus to form large chips with very complex functionality. But future generations of systems-on-chip (SoC) will consist of hundreds of pre-designed IPs assembled together to form large chips with very complex functionality. As technology scales and chip integrity grows, on-chip communication is playing an increasingly dominant role in System-on-Chip (SoC) design. To deal with the increasingly difficult problem of on-Chip communication, it has been recently proposed to connect the IPs using a Network-on-Chip (NoC) architecture. In NOC each core is connected to a switch by a network interface.

Cores communicate with each other by sending packets via a path consisting of a series of switches and inter-switch links. Network-on-Chip (NoC) has been proposed as a solution to provide better modularity, scalability, reliability and higher bandwidth compared to bus-based communication infrastructures.[2,3,4]

Fig. 1 shows an abstract view of a NOC in this architecture. As shown in fig.1, A typical NoC consists of four major components: Cores (C), Network Interface (NI) Units, Switches (S) and Physical Links. Each core can be a processing Element (PE), embedded memory, DSP or etc. Other components constitute the communication fabric.The router is connected to the four neighboring tiles and its local resource via channels. Each channel consists of two directional point-to-point links between two routers or a router and a local resource [5, 6, 7, 9].

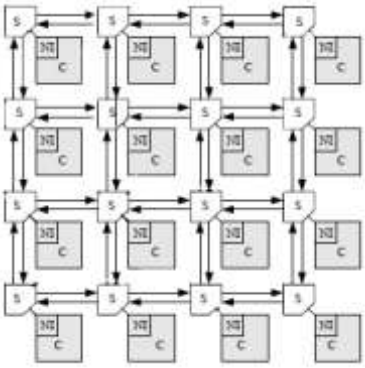

Figure 1: Abstract view of a 4*4 2-Dmesh-based NOC.Cores or (C), Network Interface (NI) Units, Switches (S)

The problem of defining communication protocols for these NoCs is not an easy matter since the resources used in traditional networks are not available on-chip. The performance of NoC largely depends on the underlying routing technique, which chooses a path for a packet and decides the routing behavior of the switches. Routing algorithms can be generally classified into two types: deterministic and adaptive. In deterministic routing, the path is completely determined by the source and the destination address. On the other hand, a routing technique is called adaptive if, given a source and a destination

address, the path taken by a particular packet depends on dynamic network conditions (e.g. congested links due to traffic variability)[6,7,8,9,10]. A routing technique has two constituencies: output selection and input selection. A packet coming from an input channel may have a choice of multiple output channels. The output selection chooses one of the multiple output channels to deliver the packet. Similarly, multiple input channels may request simultaneously the access of the same output channel; the input selection chooses one of the multiple input channels to get the access [15]. Almost all of the researches on routing techniques for NoC have focused on the improvement of output selection. Two input selection methods have been used in NoC, First-Come-First-Served (FCFS) input selection and Round-Robin input selection.

In [15], this paper investigates the impact of input selection, and presents a contention-aware input selection (CAIS) technique for NoC that improves the routing efficiency. ut there is some problem and defection in this technique. The motivation of this paper is to improve this input selection technique to develop a simple yet effective input selection technique.

The reminder of this paper is organized as follows. In Section 2, we review the related work. In Section 3, we present the proposed improved contention-aware input selection technique (ICAIS). The simulation results are presented in Section 4. Finally, Section 5 contains the summary and the conclusion of the paper.

2. Related Works

Routing strategies have a key role on communication andperformance in on-chip interconnection networks.several efforts have been done attempting to improve the performance of them in on-chip interconnection networks. In [7], a partially adaptive routing algorithm, called turn model which is based on prohibiting certain turns during routing packets to prevent deadlock is presented. In [8] a routing algorithm called odd-even was proposed based on turn model. It restricts some locations where turn can be taken so that deadlock can be avoided. In comparison with previous methods, the degree of routing adaptiveness provided by the model is more even for different source destination pairs. A routing scheme called DyAD was proposed in [9]. This algorithm is the combination of a deterministic routing algorithm and an adaptive routing algorithm. The router can switch between these two routing modes based on the network's congestion. Another adaptive routing named DyXY along with an analytical model based

on queuing theory for a 2D mesh has been proposed [10]. The authors claim that DyXY ensured deadlock-free and livelock-free routing and it can achieve better performance compared with static XY routing and odd-even routing. In [11], [12] some fully fault tolerant routing algorithms are explained, one of them is named directed flooding algorithm. In this algorithm a message is sent to each outgoing link with probability p which is not fixed but varies based on the destination of the packet. In [13] a source routing algorithm called Predominant Routing was proposed which exploits the advantages of both deterministic and adaptive routing algorithms. Also in [14] a routing algorithm for avoiding congested areas using a fuzzy-based routing decision is proposed.

All of these routing techniques focused on the output selection. Two input selections have been used in NoC, First-Come-First-Served (FCFS) input selection and Round-Robin input selection. In FCFS, the priority of accessing the output channel is granted to the input channel which requested the earliest. Round-robin assigns priority to each input channel in equal portions on a rotating basis. FCFS and Round-robin are fair to all channels but do not consider the actual traffic condition. In [15], this paper investigates the impact of input selection, and presents a novel contention-aware input selection (CAIS) technique for NoC that improves the routing efficiency. When there are contentions of multiple input channels competing for the same output channel, CAIS decides which input channel obtains the access depending on the contention level of the upstream switches, which in turn removes possible network congestion. but there is a starvation possibility in this technique. The motivation of this paper is to improve input selection to develop a simple yet effective input selection technique.

3. Improved Contention Aware Input Selection technique (ICAIS)

Each routing algorithm has two constituencies: output selection and input Selection. In this section we will present a novel input selection technique for NOC. The proposed input selection technique can be combined with an output selection, either deterministic or adaptive, to complete the routing function.

In this paper, the XY routing [8] is used as a representative of deterministic output selection for its simplicity and popularity in NoC. To avoid deadlock, the minimal odd-even (OE) routing [8] is used as a representative of adaptive output selection.

Multiple input channels may request simultaneously the access of the same output channel, e.g., packets p0 of input_0 and p1 of

input_1 can request output_0 at the same time. The input selection chooses one of the multiple input channels to get the access. Two input selections have been used in NoC, First-Come-First-Served (FCFS) input selection and Round-Robin input selection. In FCFS, the priority of accessing the output channel is granted to the input channel which requested the earliest. Round-Robin assigns priority to each input channel in equal portions on a rotating basis. FCFS and Round-Robin are fair to all channels but do not consider the actual traffic condition. This section presents an input selection that performs more intelligent, by considering the actual traffic condition, leading to higher routing efficiency. In this paper we consider NoCs with 2D mesh topology . Wormhole switching is employed because of its low latency and low buffer requirement. Similar to [15], the basic idea is to give the input channels different priorities of accessing the output channels. The priorities are decided dynamically at run-time, based on the actual traffic conditions of the upstream switches. More precisely, each output channel within a switch observes the contention level (CL)(the number of requests from the input channels) and sends this contention level to the input channel of the downstream switch, where the contention level is then used in the input selection. When multiple input channels request the same output channel, the access is granted to the input channel which has the highest contention level acquired from the upstream switch. Fig 2 shows the concept of Contention Level(CL) in a switch. This input selection removes possible network congestion by keeping the traffic flowing even in the paths with heavy traffic load, which in turn improves routing performance. Such an input selection helps reduce the number of waiting packets in congested areas. This removes possible network congestions and leads to better NoC performance. Based on this observation a input selection is developed. For the input channels connected to the cores, there are no upstream switches transmitting CL to them. The CL value is set to 0 for these input channels. Therefore, the packets already in the network have higher priority than the packets waiting to be injected into the network.

With a little attention to above input selection technique (CAIS), we notice an important problem. If an input channel which has lower CL continuously competing with channels which have higher CL, obviously will be defeated any time. The packets in this channel won't be able to get their required output channel and face with starvation and this will cause the problem of decreasing network

efficiency. Thus, there is a starvation possibility in this input selection technique, because it performs input selection only based on the highest contention level (CL) and the channels with low CL have a little chance for winning. So, now we try to consider priority parameter in a way that input channels with low CL, have the opportunity to win. Therefore, in addition to CL, another parameter with the name of AGE for every input channel is taken into consideration and measure of priority will be a compound of CL+AGE (Fig 3).

The initial value of AGE for every channel is zero. When some input channels compete each other to achieve a specific output channel, finally only one channel will succeed.

After this, the AGE of the winner channel will reset to zero and AGE of the other channels entering in the competition will increase for one unit. With this new criteria(CL+AGE) each time that an input channel compete with other input channels to achieve specific output channel, in case of failure, it's AGE increase one unit and this increase its priority for the following competitions. This itself increase the opportunity of success and finally this channel be able to gain its desired output channel. In competition the following conditions may occurs:

a) if the priority(CL+AGE) of an input channel be higher than of other input channels, then the desired output channel will be granted to it and then its AGE will reset to zero. Then the AGE of all other input channel will increase for one unit.

b) If multiple output channels have the equal priority, the output channel will be granted to that input channel which has higher AGE. Then its AGE will be reset to zero and the AGE of other input channels will increase for one unit.

Fig 4 shows Pseudo code of ICAIS input selection technique.

Fig. 5 illustrates the detailed architecture of a switch with new input selection technique(ICAIS). As can be seen, the structure of this switch is similar to CAIS switch, with slight difference. Here, a very small unit to compute the AGE parameter is added to each port.

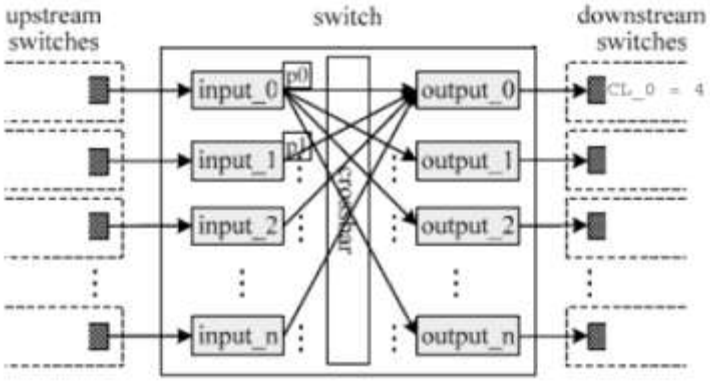

Figure 2: Concept of Contention Level(CL) in a switch

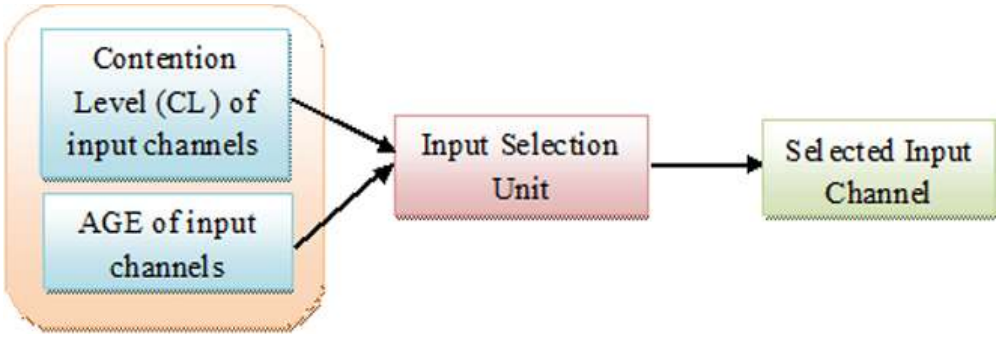

Figure 3: Priority parameters in ICAIS

```
Input selection :
IS function :  for (  ;  ;  ) {
CL= # of access request to j th output channel;
send CL to downstream switch;
for each input channel i do{   AGE i = # of unsuccessful try ;
M i = observed (CLi )+ AGE i;}
In each competition do {
 if M i > M j,Mk,... then { output channel granted to channel i;
AGE i = 0; increament AGE of M j,Mk,...}
 if M i= =M j then
   if AGE i > AGE j then { grant the access to channel i;
          AGE i = 0;     increament AGE of M j,Mk,...; }
     else if AGE i < AGE j then{
        get the access to channel j;   AGE j=0;
        increament AGE of     M i,Mk,...; }
           else   { get acces to a chanel randomly;
                winner AGE = 0;
                increament AGE of other channel.}   } }
```

Figure 4: Pseudo code of proposed input selection technique (ICAIS)

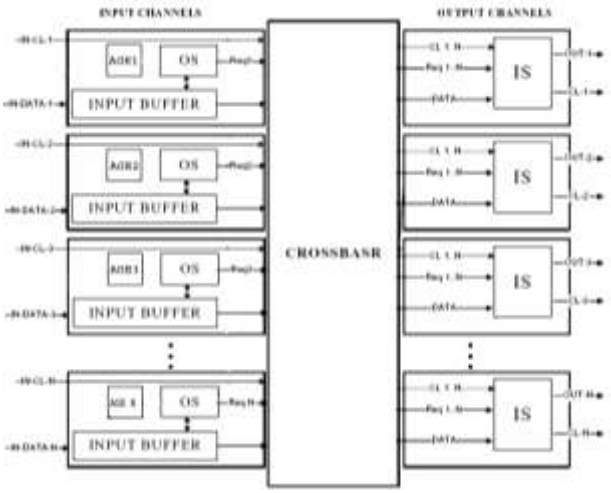

Figure 5: Switch architecture with ICAIS input selection technique

RESULTS

To evaluate the performance gains that can be achieved with our new input selection technique (ICAIS), we developed a C++ based simulator. Experiments are conducted to evaluate the performance of the ICAIS input selection technique and give a comparison between ICAIS, CAIS and traditional input selections. Due to the advancement of FCFS over round-robin, FCFS is selected to compare with CAIS. All input selection techniques are combined with a deterministic output selection (XY routing) and an adaptive output selection (OE routing). The network size during simulation is fixed to be 6*6 tiles. It is assumed that the packets have a fixed length of 5 flits and the buffer size of input channels is 5 flits. The efficiency of each type of routing is evaluated through latency-throughput curves. Similar to other work in the literature, we assume that the packet latency spans the instant when the first flit of the packet is created, to the time when last it is ejected to the destination node, including the queuing time at the source.

For each simulation, the packet latencies are averaged over 50,000 packets. Latencies are not collected for the first 5,000 cycles to allow the network to stabilize. Since the network performance is greatly influenced by the traffic pattern, in this set of experiments we consider three synthetic traffic patterns: uniform, transpose, and hot spot. In the uniform traffic pattern, a core sends a packet to any other cores with equal probability. In the transpose traffic pattern, a core at (i, j) only send packets to the core at (5-j, 5-i). In the hot spot traffic pattern, the core at (3, 3) is designated as the hot spot, which receives 10% more traffic in addition to the regular uniform traffic.

Fig. 6 shows the performance of the six routing schemes under uniform traffic. As can be seen from the figure, the four schemes have almost the same performance at low traffic load (<0.040 packets/cycle). As the traffic load increases, the packet latency rises dramatically due to the network congestion. Comparing the curves of OE+FCFS , OE+CAIS and OE+ICAIS it can be seen that, using the OE output selection, ICAIS performs better than CAIS and FCFS. Similarly, the curves of XY+FCFS, XY+CAIS and XY+ICAIS show that ICAIS also outperforms FCFS and CAIS when using XY output selection. Fig. 7 shows the performance of the six routing schemes under transpose traffic. It can be seen that FCFS and CAIS and have the same performance when using the XY output selection; FCFS works slightly better than CAIS and ICAIS when using the OE output selection. This is because with transpose traffic, it is rarely the case that more than one input channels compete for the same output channel. Therefore, the input selection policy has little impact on the routing performance. Fig. 8 shows the routing performance under hot-spot traffic. Once again, it can be seen that ICAIS significantly outperforms FCFS and CAIS, either using XY or OE output selection.

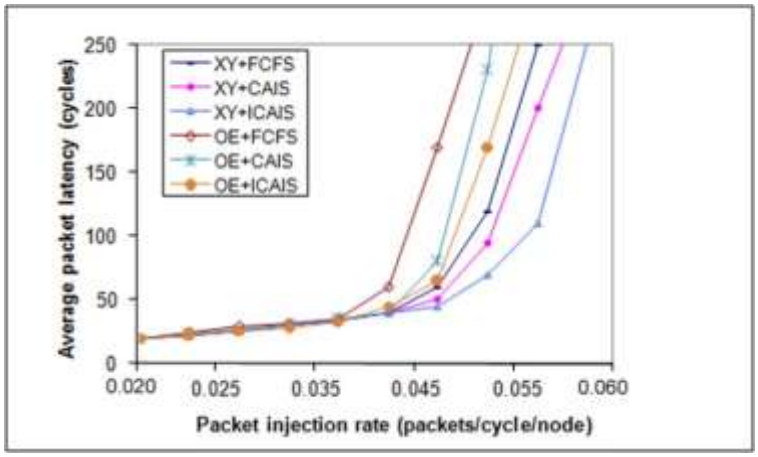

Fig 6: Performance of routing schemes under uniform traffic

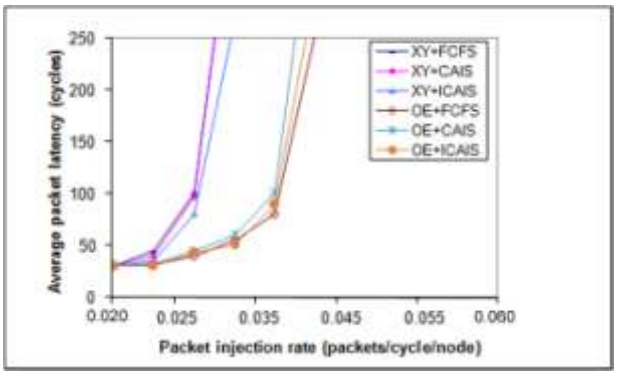

Figure 7: Performance of routing schemes under transpose traffic

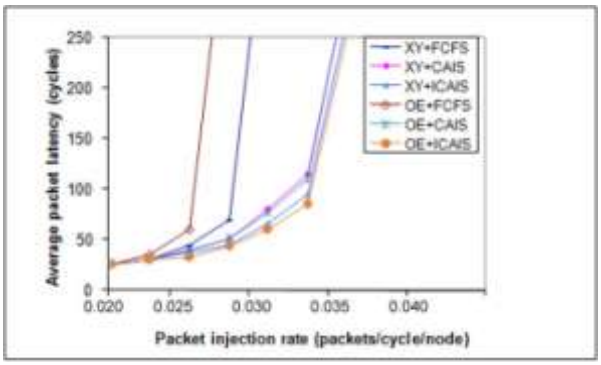

Figure 8: Performance of routing schemes under hot spot traffic

CONCLUSION

The performance of Network-on-Chip largely depends on the underlying routing techniques. A routing technique has two constituencies: output selection and input selection. This paper has shown the importance of input selection in routing efficiency. In this paper an new efficient input selection technique, ICAIS, is presented which performs more intelligent ,by considering the actual traffic condition of the network, leading to higher routing efficiency. In this paper we improve the problems and defections of contention-aware input selection (CAIS) technique to develop a simple yet effective input selection technique, ICAIS. The simulation results show the effectiveness of ICAIS by comparing it with CAIS and traditional input selections.

REFERENCES

1. A.Mehran, A., Saeidi, S., khademzadeh, A.: Spiral: A heuristic mapping algorithm for network on chip. IEICE Electronic Express, Vol.4, No.15, 478.Japan,(2007)

2. Driankov, D., Hellendoom, H., Reinfrank, M.,: An introduction to fuzzy control, Springer- Verlag, Berlin, New York.(1993)

3. Benini, L., De Micheli, G.,: Networks on-chip: a new soc paradigm. Journal of *IEEE* Computer,vol 35 : 70-78.(2002)

4. Kumar, S., Jantsch, A,.Soininen, J. P,.Forsell, M,. Millberg, M,.Oberg, J.A ., : network on chip architecture and design methodology," ISVLSI, pp. 117-24, USA, (2002)

5. Ivanov, A,.and Micheli, G. D,. : The Network -on-Chip Paradigm in Practice and Research. IEEE Design and Test of Computers, vol. 22, no. 5, pp. 399-403, (2005)

6. Behrouzian-Nejad, E., Khademzadeh, A., : BIOS: A New Efficient Routing Algorithm for Network on Chip", journal of Contemporary Engineering Sciences, Vol. 2,no. 1, 37 – 6.(2009)

7.Glass, C. J., Ni. L. M.,: The Turn Model for Adaptive Routing. In ISCA '92: Proceeding of the 19th annual international symposium on Computer architecture.(1992)

8. Chiu, G. M.,: The Odd-Even Turn Model for Adaptive Routing.IEEE Trans. on Parall. And. Dist. Sys.1: no.7 (2002)

9. Hu, J., Marculescu, R.,: DyAD-Smart Routing for Networks-on-Chip. Proceeding of DAC 2004, San Diego,California,USA. pp. 260 – 263.(2004)

10. Li, M., Zeng, Q. A., Jone, W. B.: DyXY-A Proximity Congestion-Aware Deadlock-Free Dynamic Routing Method for Networks-on-Chip. : Proceedings of ACMlIEEE Design Automation Conf., pp. 849-852.(2006)

11. Pirretti, M., Link, G. M., Brooks, R. R., Vijaykrishnan, N., Kandemir, N. M., Irwin, M. J., :Fault tolerant algorithms for network-on-chip interconnect. Proceeding of IEEE Computer society Annual Symposium on 19-20, pp. 46 – 51.(2004)

12. Dumitras, T., Kerner, S., Marculescu, R., : Towards on-chip fault tolerant communication. Proceedings of Asia and South Pacific Design Automation Conference.(2003)

13. Asad, A., Seyrafi, M., Ehsani Zonouz, A., Seyrafi, M., Soryani M., Fathy, M.: A Predominant Routing for On-Chip Networks. Proceeding of IDT, Riyadh, pp. 1-6.(2009)

14. Salehi, N., Dana, A., : A fuzzy-based power-aware routing algorithm for network on-chip. Proceeding of ICACT 2010, Phoenix Park, pp 1159-1163.(2010)

15. Wu, D., Al-Hashimi, B. M., and Schmitz, M. T.: Improving Routing Efficiency for Network-on-Chip through Contention-Aware Input Selection.Proceeding of 11th Conference (ASP-DAC 2006).