Age-aware Adaptive Routing for Network-on-Chip Routing with Odd-Even Turn Model

M. S Saliu¹, I. J Umoh², B.O. Sadiq² and M. O Momoh²

¹Basic Sciences Department, Federal College of Freshwater Fisheries Technology, New Bussa, Nigeria
²Department of Computer Engineering, Ahmadu Bello University Zaria, Nigeria

E-mail address: smshaba@abu.edu.ng

Abstract—This paper presents an age-aware adaptive routing for Odd-Even (OE) turn model. As packets traverse from source to destination node, their paths are defined by a given routing algorithm. For a selected routing algorithm, an efficient arbitration technique is crucial to sharing critical Network-on-Chip resources. Arbitration techniques provide high degree of local fairness from each router point of view. However, there is delay of a packet with a longer path between the source and destination nodes. In order to address this challenge an age-based arbitration technique is hereby proposed for adaptive routing with OE turn model. The age-aware adaptive routing uses an age-based arbitration technique that gives priority to oldest packet. The performance of the developed age-aware adaptive routing was evaluated using different synthetic traffic at different Packet Injection Rates (PIRs). Results were compared with the result obtained on fair arbitration technique for adaptive routing using average latency and throughput as performance metrics. The result indicated that the age-aware adaptive routing has 2.73%, 6.63 %, 5.4% and 4.5 % reduction in latency under random, transpose 1 transpose 2 and bit reversal traffic patterns respectively when compared to fair arbitration adaptive routing with OE turn model. For throughput the results indicated that the age-aware adaptive routing with OE turn model has 14.22%, 13%.12% and 19% increase in throughput under random, transpose 1 transpose 2 and bit reversal traffic patterns respectively when compared to fair arbitration adaptive routing with OE turn model .

Keywords - Network-on-Chip; Adaptive routing; Odd-Even turn model; Arbitration Technique; Arbiter; Age-Based Arbitration; Average latency; Throughput

I. INTRODUCTION

In the state-of-the-art SoC designs, the foremost problem is the interconnection between the components- IP cores or PEs ([1], [2]). Network-on-Chip (NoC) has been proposed as a viable solution to the communication challenges on SoCs ([3], [4]).

Routing algorithms determine the path a packet follows from the source PEs to the destination PEs ([5],[6]). For a classical deterministic routing method such as XY routing, a given packet is transferred from the source PE to the destination PE through a specific predetermined path irrespective of traffic conditions. XY routing is deadlock free and relatively simple to implement with low power consumption [7]. However, as the number of nodes or PEs increases, it is not able to avoid congested regions in a busy and non-uniform traffic. Thus, affecting the throughput and network latency. For an improved network performance, a routing algorithm with an alternate path to transfer packets is required [8].

In an adaptive routing algorithm, the path for packet transversal from source to destination is dynamically selected based on the condition of the network thereby avoiding congested regions. Several studies have shown that classical fully adaptive routing algorithm is prone to deadlock without Virtual Channels (VCs) ([9],[10],[11]). However, the addition of VCs comes with a significant increment in hardware required and complexity in the design of the routers, which potentially brings about increased hardware overhead, power consumption, and network latency [12]. To minimize the aforementioned problems, turn model was proposed by [13]. It is based on the turn model that three partially adaptive routing algorithms were proposed, namely: west-first, negative-first, and north-last. Basically, in turn model, a turn is 90º change in the direction of a packet and there can either be clockwise or counterclockwise turn. Therefore, deadlock can be avoided if some turns are prohibited [13]. But with turn prohibition based on global routing, the adaptiveness becomes highly uneven, some path between the source node and destination node have minimal adaptiveness while the remaining paths are fully adaptive. The performance of NoC may be affected due to this uneven adaptiveness because it limits the ability of the turn model to improve the congestion problem [14]. In alleviating this problem, [15] proposed a turn model called OE turn model that prohibits turns based on the column the node is located instead of globally prohibiting certain turns.
Additionally, the OE turn model is reported to have more paths to route packets than other turn models [16].

For a conventional OE turn model as odd and even columns appear alternatively, the model is limited to a relatively determined path. To improve candidate path, balance network traffic and minimize congestion in NoC, [17] employed a fair alterable priority arbitration technique as the selection strategy. The packet traversal is determined by the arbitration granting result for all the permitted path by OE turn model. The alterable priority arbitration technique uses allocation principle of fixed priority arbiter and an optimized pointer scheme similar to pointer scheme of Round Robin (RR) arbiter to change the priority order. The conventional RR and other improved RR arbitration technique [17] typically used in NoC router provides a strong fairness from a single router point of view i.e local fairness. However, since packets traverses several routers on their way to their destination nodes, fairness is basically a global property as such a mechanism for coordinating the individual routers to allocate output port based on global fairness is required.

The contribution of this paper is the proposal of an age-aware adaptive routing algorithm with OE turn model. The latency and throughput were evaluated.

This paper is structured as follows; In section II, we presented the previous researches on fairness of arbitration techniques. In section III, the age-aware adaptive routing technique was presented. The result analysis which includes simulation setup and simulation results are presented in section IV. Finally, Section VII gives conclusion as well as future research directions.

II. RELATED RESEARCH

There has been several techniques and combination of techniques used to improve fairness of arbitration technique in NoC but most of it suffer from increased implementation complexity and hardware overhead. In [18] to improve fairness of the arbitration technique the authors implemented arbitration technique which assign weight based on the number of contending flow at each input port of the router. The weight depends on the size of the time frame. Increasing the time frame reduces frequency of updating the weight while reduction in the time frame causes limitation in maximum weight assigned. To assign accurate weight all source nodes sharing the same path sends at least one packet. In implementing the technique source address is required for each flit, the module tracks the source addresses which increases linearly as number of nodes increases in the system which in turn increases it complexity of implementation.

In the work of [19] an arbitration technique that is priority-based was implemented. In the technique, multiple Time-Slots (TSS) is allocated to each VC locally. Some of the drawbacks of this method include; increased overhead cost due to the implementation of the TSS, the TSS significantly reduces the operating frequency of the routers and generally the TSS are expensive to implement. In [20] the authors implemented an age-based priority arbitration technique. A packet that has stayed long in the network is given higher priority. With the arbitration there was improved fairness as indicated in their results obtained but there is increased cost of complexity in maintaining the age of each flits. Authors in [21] on the other hand implemented an arbitration technique with priority based on the distance travelled by a given packet which is similar to the work of [20]. To assign weight to the packets several methods was used (from fixed to dynamic). The packets with higher weight have higher probability to being forwarded first. However, in each clock cycle, a bounded random number is generated for probabilistic Weighted Round-Robin Arbiter (WRRA). This incurs a significant increase in area overhead and latency. To minimize the latency the authors later implemented an architecture that pre calculates the random numbers a clock in advance. To further minimize the latency, a Round-Robin Arbiter (RRA) combined with probabilistic arbitration technique determine the arbitration result if the precalculated result is not ready. As indicated by the result they achieved reduction in latency but there was significant increase in area overhead. In [22] implemented another mechanism that adaptively calculates weight based on remaining hop count instead of traversed ones. The modified flow control implemented changes priority among different packets based on network status. The probabilistic arbitration is expensive with respect to hardware overhead. moreover, using the remaining hop count as global fairness metric is not accurate. In [17] proposed an alterable priority arbitration technique for fair granting of packet transmission from all the possible directions permitted by OE turn model. The alterable priority arbitration technique use granting mechanism of a fixed priority (FP) and it fairly varies the priority order when the optimized pointer value increases by 1. By applying the proposed arbitration technique with OE turn model, the network latency was reduced and the capacity of the network was improved. The area overhead of the proposed routing algorithms was also evaluated and was found to have minimal increase in area overhead when compared to the selected routing algorithms. But as more packet get injected into the network and the network load increases the Network-on-Chip gets congested and consequently saturated.

III. AGE-AWARE ADAPTIVE ROUTING TECHNIQUE

The routing efficiency of NoC is dependent on the efficiency of the port selection technique employed on the router. Basically, the port selection technique includes input-selection technique and output-selection technique. The technique employed for output-selection determines the output port that the flit is delivered. Input-selection technique on the other hand selects one of the input channels to be granted access to a given output port. For example, flit f0 of input_0 and flit f1 of input_1 requesting output_1 the same time. Two types of input-selection techniques mostly used in NoC architecture, First-Come-First-Served (FCFS) and RR input-selection arbitration techniques. Both techniques do not consider priority of the flits as one of the input ports is granted access to a given output port. Each of the aforementioned arbitration techniques are fair to each input port but the fairness is only limited from local ports point of view. But flit might have undergone several contentions for network
resources on it path from source to destination and need to be prioritized when competing for network resources at a given port thereby improving the overall global fairness of the network.

To improve the global fairness and improve the overall performance, this research proposes an age-aware adaptive routing with OE turn model that implements an output-selection technique using alterable priority arbitration technique – an improved RR by [17] and an age-based arbitration technique for the input-selection technique. For the proposed age-based arbitration technique, if more than one input port attempt to access an output port, the arbitration technique compares the ages of flits vying for the same output port. The arbiter then selects the input port whose packet has stayed longest in the network. In a situation in which two or more packets has the same age, the arbiter randomly selects one of the competing input ports. The algorithm for the age-based arbitration technique is listed in Algorithm.

Algorithm 1: Proposed Age-Based arbitration

```c
for ( j = 0; j < CL; j++) { 
  for (channel = i; channel < MAX_CHANNELS; channel++) 
     AGEj = no. of lost contentions Mj = AGEj; 
  
  for each competition in 
     if Mj > Mi, Mj,... then 
        output channel granted to channel 
        AGEj = 0; 
        increment age of Mj; 
        if AGEj = AGEi then 
            grant the access to 
            AGEj = 0; 
            increment age of AGEi; 
        else if AGEj < AGEi then 
            grant the access to channel 
            AGEj = 0; 
            increment age of Mj; 
        else 
            grant the access to random channel 
            AGETj = 0; 
            increment age of other channels; 
  
  WinNER = channel = i; channel < MAX_CHANNELS; 
  ```
```

IV. RESULT ANALYSIS

A. Simulation Setup

The feasibility of the proposed age-based priority arbitration scheme was evaluated using an improved cycle-accurate NoC simulator Noxim. The performance of the proposed age-aware adaptive routing with OE turn model and fair arbitration adaptive routing with OE turn model was evaluated using the Noxim simulator. The Noxim simulator is based on System C of Electric System Level(ESL)[25]. The NoC configuration parameters used for the simulation is listed in Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network topology</td>
<td>4x4 Mesh</td>
</tr>
</tbody>
</table>

To analyze the influence of the age-based arbitration technique on the network performance, the performance metrics of adaptive routing based on OE turn model applying fair arbitration technique as selection strategy [17] and OE applying the proposed age-based arbitration technique as input port selection strategy are compared in 4X4 2D mesh.

To verify the proposed routing algorithm, the two important performance metrics of average latency and throughput are evaluated using different traffic patterns. The average latency is calculated as average time instant it takes flits to reach their destination nodes from the time instant, they are created in network which includes the queuing time at source nodes while throughput measures the fraction of the maximum load a network is capable of handling. The formula for average latency and throughput is expressed in equations 1 and 2 respectively [23].

\[
T_p = \frac{T_f}{N \times T_c} \tag{1}
\]

\[
T_{average} = \frac{1}{N} \sum_{i=1}^{N} L_i \tag{2}
\]

With K representing the total number of packets that reaches a given destination node, L is equivalent to the delay of each packet i . T represents Total received flits, N represents number of nodes and Tc denotes Total cycles.

The spatial distribution of packets delivered from source node to destination nodes in the network is determined by the traffic patterns, random (RD), transpose 1(TR1), transpose 2(TR2) and bit reversal (BR) traffic patterns are supported by Noxim simulator [24]. In random traffic pattern, packets are sent by each router to other routers in a random manner. The pair of source and destination nodes are also selected randomly. Transpose traffic pattern can either be transpose 1 or transpose 2. In transpose 1, packet from a source router with address (i,j) would only be sent to router of destination address (N-1-j,N-1-i) where N is given as the mesh size. In transpose 2 packet from source (i,j) would only be sent to destination address (j,i).With bit reversal, the packets from the source node are sent to destination address which is bit reversed of the source address. Distributing the source node (i,j) onto the entire. With source node (i,j) distributed onto the entire network, the performance metrics of routing algorithms can be simulated according to a specific traffic patterns.

B. Simulation Results

As an important performance metric, average latency is used to evaluate the proposed routing algorithm. The packet latencies of OE turn model adaptive routing applying fair
arbitration as output selection strategy and OE turn model applying age-based arbitration technique as input selection strategy is illustrated in Fig 1 using different traffic scenarios, with a packet size of 3 flits in a 4x4 2D mesh topology. The proposed age-aware adaptive routing with OE turn model was compared to the work of [17] to evaluate the performance of the proposed routing algorithm due to reduce hardware overhead of both routing algorithm. It can be observed that the fair arbitration adaptive routing with OE turn model [17] performs slightly better in terms of average latency than age-aware adaptive routing with OE turn model at lower PIR, this is due to higher complexity in determination of path by flits which incurs more delay. But as the PIR increases the age-aware adaptive routing with OE turn model performs better in terms of average latency which is an indication that the network speed has improved by employing the proposed age-based arbitration as input selection strategy. Since as more flits get injected into the network fair alterable priority arbitration technique is oblivious of the number of contentions a flit has experienced, it can end up granting a flit that has undergone little or no contention for network resources while delaying a flit that has undergone several contentions for network resources thereby increasing it latency and by extension the overall average latency of the network. The age-aware adaptive routing with OE turn model has 2.73%, 6.63%, 5.4% and 4.5% reduction in latency under random, transpose 1 transpose 2 and bit reversal traffic patterns respectively when compared to fair arbitration adaptive routing with OE turn model as shown in Table 2. By prioritizing the oldest flit first, the age-aware arbitration technique firstly increases the age of a flit increases with increase contention for network resources thereby minimizing the likelihood of incurring increasing delay as flit transverses from source node to destination node. Secondly, the age-aware arbitration technique prioritizes the flit with oldest age, hence providing global fairness, generally reducing the variance in flit transient time and overall average latency of the NoC was reduced.

The throughput is another important performance metric to measure the capacity of NoC, the capacity of a network can be improved by making arbitration technique globally fair in a network, which reduces the variance in throughput among different packet flows. Fig 2 illustrates the throughput of OE turn model adaptive routing applying fair arbitration as selection strategy and OE turn model adaptive routing applying age-base arbitration technique as input selection strategy under different traffic scenarios. It can be observed that for all the traffic patterns, both age-aware adaptive routing with OE turn model and fair arbitration adaptive routing with OE turn model [17] exhibited similar degree of linear increase in throughput which indicate that at lower PIR there is lower contention for network resources hence similar number of packets get delivered from source to destination node in both routing techniques. However, as the throughput raises steadily it gets to certain throughput value and the network throughput begins to drop as shown in the Fig 2.

<table>
<thead>
<tr>
<th>S/ N.</th>
<th>PIR(Packet/ Cycle/node)</th>
<th>RD</th>
<th>TR 1</th>
<th>TR 2</th>
<th>BR</th>
<th>RD</th>
<th>TR 1</th>
<th>TR 2</th>
<th>BR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.10</td>
<td>14.85</td>
<td>11.90</td>
<td>11.99</td>
<td>11.04</td>
<td>14.53</td>
<td>11.20</td>
<td>11.56</td>
<td>10.33</td>
</tr>
<tr>
<td>2</td>
<td>0.15</td>
<td>16.62</td>
<td>12.72</td>
<td>12.71</td>
<td>11.07</td>
<td>16.00</td>
<td>12.60</td>
<td>12.50</td>
<td>10.83</td>
</tr>
<tr>
<td>3</td>
<td>0.20</td>
<td>19.47</td>
<td>13.04</td>
<td>14.10</td>
<td>12.05</td>
<td>19.33</td>
<td>13.30</td>
<td>14.00</td>
<td>12.16</td>
</tr>
<tr>
<td>4</td>
<td>0.25</td>
<td>20.10</td>
<td>13.70</td>
<td>14.45</td>
<td>13.90</td>
<td>21.41</td>
<td>14.66</td>
<td>15.60</td>
<td>15.00</td>
</tr>
<tr>
<td>5</td>
<td>0.30</td>
<td>24.90</td>
<td>16.20</td>
<td>16.47</td>
<td>17.28</td>
<td>26.33</td>
<td>18.60</td>
<td>18.00</td>
<td>18.83</td>
</tr>
<tr>
<td>6</td>
<td>0.35</td>
<td>27.00</td>
<td>18.90</td>
<td>19.25</td>
<td>19.96</td>
<td>28.50</td>
<td>21.50</td>
<td>21.30</td>
<td>22.00</td>
</tr>
<tr>
<td>7</td>
<td>0.40</td>
<td>27.99</td>
<td>21.70</td>
<td>22.90</td>
<td>22.28</td>
<td>30.40</td>
<td>25.30</td>
<td>26.00</td>
<td>24.60</td>
</tr>
<tr>
<td>8</td>
<td>0.45</td>
<td>29.82</td>
<td>24.30</td>
<td>24.30</td>
<td>24.80</td>
<td>31.62</td>
<td>27.80</td>
<td>27.67</td>
<td>27.60</td>
</tr>
</tbody>
</table>

Table 2: Average latency for Age-aware adaptive routing and Fair arbitration adaptive routing at different traffic patterns at different PIR.
Even turn model that uses globally fair age-based arbitration technique to select output port for lower average latency and high throughput. This improves the overall performance of the Network-on-Chip by reducing the average latency by 4.8% and increasing the throughput by 14% when compared fair arbitration adaptive routing. The comparison of the performance of age-aware adaptive routing with Odd-Even turn model and fair arbitration adaptive routing with Odd-Even turn model were carried out for average latency and throughput under different traffic patterns. The age-aware adaptive routing outperformed the fair arbitration adaptive routing for all the performance metrics.

The research only considered synthetic traffics for performance evaluation, real media applications can be used for evaluating the performance of the proposed adaptive routing in future works.

REFERENCES


---

Table 3: Throughput for Age-aware adaptive routing and Fair arbitration adaptive routing at different traffic patterns at different PIR

<table>
<thead>
<tr>
<th>S/N</th>
<th>PIR(Packet/Cycle/Node)</th>
<th>Age-aware adaptive routing</th>
<th>Fair arbitration adaptive routing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.10</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>2</td>
<td>0.15</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>3</td>
<td>0.20</td>
<td>0.21</td>
<td>0.19</td>
</tr>
<tr>
<td>4</td>
<td>0.25</td>
<td>0.32</td>
<td>0.22</td>
</tr>
<tr>
<td>5</td>
<td>0.30</td>
<td>0.38</td>
<td>0.39</td>
</tr>
<tr>
<td>6</td>
<td>0.35</td>
<td>0.37</td>
<td>0.36</td>
</tr>
<tr>
<td>7</td>
<td>0.40</td>
<td>0.36</td>
<td>0.37</td>
</tr>
<tr>
<td>8</td>
<td>0.45</td>
<td>0.35</td>
<td>0.36</td>
</tr>
</tbody>
</table>

This maximum throughput is indicative of the capacity of the routing algorithm which is also known as the saturation throughput. From Fig. 2 the saturation throughput under fair arbitration adaptive routing with OE turn is 0.34, 0.37,0.38 and 0.327 for random, transpose 1 transpose 2 and bit reversal respectively compared to 0.38,0.39,0.4 and 0.38 for random, transpose 1, transpose 2 and bit reversal respectively under age-aware adaptive routing with OE turn model as indicated in table 1 and 2. The age-aware adaptive routing with OE turn model has 14.22%, 13%,12% and 19% increase in throughput under random, transpose 1 transpose 2 and bit reversal traffic patterns respectively when compared to fair arbitration adaptive routing with OE turn model as shown in table 2 which indicate that age-aware adaptive routing outperforms fair adaptive routing in terms of overall throughput. This is as a result of uneven distribution of throughput among different traffic flows caused by local fairness of the fair arbitration technique. But age aware arbitration technique prioritizing of older flits corrects this imbalance among different flows. Since as a traffic flow from a given flit receives lesser share of the overall throughput the flit undergoes additional delay thereby increasing it age which in turn increases it priority. Therefore, there is general improvement in the throughput as illustrated in Fig. 2

V. Conclusion

In this paper, to minimize queuing delay and variance in throughput among different traffic flows, an age-based arbitration technique is used for the permitted turns in OE turn model. We proposed an age-based arbitration scheme input selection according to priority mechanism of oldest flits first. By applying the proposed arbitration technique for input selection, the queuing delay is reduced and network capacity improved, which resulted in better performance metrics. The simulation results under the Noxim simulator show that the introduction of the proposed arbitration technique reduces average packet latency and overall throughput compared to fair arbitration OE turn model based adaptive routing. The research developed an age-aware adaptive routing with Odd-


Ime Jarlath Umoh received the B.Eng degree in Electrical Engineering from Ahmadu Bello University, Zaria, Nigeria, in 2006, and the M.Sc. degree in microelectronics system design from the University of Southampton, U.K., in 2009 and He Ph.D. degree at the University of Southampton in 2014. In 2010 he joined the Department of Computer Engineering, Ahmadu Bello University, Zaria where he continues his research in modeling, simulation and synthesis of nano-transistors and their related applications.

His research interests include System-on-Chip(SoC), Network-on-Chip(NoC), FPGA, Machine Learning and Computer Vision.

Bashir Olaniyi Sadiq is a Lecturer of the Department of Computer Engineering, Ahmadu Bello University Zaria. He obtained his B. Eng., MSc and Ph.D. degrees in Computer Engineering from Ahmadu Bello University Zaria. He is a member of IEEE, IAENG, NSE, and Council for the Regulation of Engineering in Nigeria. His areas of Expertise are Networks and Image Processing.

Muyideen Omuya Momoh received a B. Eng. in Computer Engineering from Federal University of Technology, Minna, Nigeria, in 2014 and MSc. in Computer Engineering from Ahmadu Bello University, Zaria, Nigeria, in 2018 and currently pursuing his PhD in same University. He is a Lecturer at the Department of Aerospace Engineering, Air Force Institute of Technology, Kaduna, Nigeria. His research interests include Computer Networks, Image and Video Processing and Avionics.

Salii Mohamed Shaba received a B. Eng. in Electrical and Computer Engineering from Federal University of Technology, Minna, Nigeria, in 2007 and M. Sc. in Computer Engineering from Ahmadu Bello University, Zaria, Nigeria, in 2020. Currently, he is a lecturer at the Department of Basic Sciences, Federal College of Freshwater Fisheries Technology, New Bussa, Nigeria.