Microprocessors and Microsystems 45 (2016) 81–94

Contents lists available at ScienceDirect

Microprocessors and Microsystems

journal homepage: www.elsevier.com/locate/micpro

Improving the Area Efficiency of ACO-Based Routing by Directional Pheromone in Large-Scale NoCs

Yuhai Li[a,b], Kuizhi Mei[a,*], Yuehu Liu[a]

[a] Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, P. R. China
[b] Science and Technology on Electro-Optical Information Security Control Laboratory, Sanhe, Hebei 065201, P. R. China

ARTICLE INFO

Article history:
Received 22 July 2015
Revised 27 March 2016
Accepted 1 April 2016
Available online 4 April 2016

Keywords:
Ant Colony Optimization
Adaptive Routing
Network-on-Chip
Directional Pheromone

ABSTRACT

Ant Colony Optimization (ACO) is a distributed collective-intelligence algorithm. Several adaptive routing algorithms based on ACO have been proposed in the domain of Network-on-Chip (NoC) design for balancing traffic load. However, when network size becomes large, the conventional ACO requires quite a lot of pheromones for predicting network load distribution, which results in large hardware cost and low cost-efficiency. In this paper, an ACO algorithm with directional pheromone (ACO-DP) is proposed for reducing the size of pheromone table in large-scale networks. Moreover, by using a distance-sensitive backward pheromone updating scheme, the performance of ACO-DP is also improved. Finally, we introduce the detailed architecture and hardware implementation of ACO-DP routing. Experimental results show that ACO-DP routing achieves the highest area efficiency in large-scale NoC systems compared to other ACO-based routing algorithms.

1. Introduction

In network-on-chip (NoC) systems, cores need to communicate with each other during parallel processing [1,2]. As the number of cores scales up to tens or hundreds, the quantity of data transmission grows rapidly. Consequently, in a large-scale NoC, system performance will be more sensitive to transmission efficiency than a small-scale one [3,4]. In NoC systems, transmission performance is mainly affected by packet routing. Routing algorithms are classified into deterministic routing and adaptive routing [2,5]. When using a deterministic routing, the transmission paths are fixed, so load-balancing is weak and passive especially for large-scale networks [6]. In contrast, adaptive routing can dynamically determine the output channels and actively find an appropriate path for each packet based on network load. Therefore, to minimize traffic congestion and improve transmission efficiency, the large-scale NoC systems should adopt an effective adaptive routing algorithm for load balancing [7–9].

Typically, an adaptive routing is composed of routing module and selection module [10]. The routing module computes several candidate output ports based on turn models, such as Odd-Even [11] and DyAD [12]. Then the selection module grades those candidate output channels and chooses one according to buffer utilization rate or predicted load state. Obviously, the selection module is the key point of an adaptive routing, because its selection scheme mainly determines packets' transmission path and gives a significant impact on network load distribution. Therefore, designing an effective selection scheme can improve transmission performance for all adaptive routing algorithms in NoC systems.

Considering the hardware cost of router, most adaptive routing algorithms determine the transmission path based on the locality of network states. It works well in a small-scale network, but when the network size scales up, using regional traffic load to choose the transmission paths may result in global imbalance. Besides, congestion estimates based only on regional network information is not sufficiently accurate [13]. So, in the large-scale network, adaptive routing should collect both regional and global network states to select the output channel for each packet during data transmissions. Because a routing table for saving global network states is very large, some adaptive routing algorithms with learning and training are proposed, such as routing algorithms based on ACO (Ant Colony Optimization).

ACO is a distributed collective-intelligence algorithm which inspired from the foraging behavior of real ant colonies [14]. During searching food, ants deposit pheromone on the ground in order to mark their routes, also they follow other ants according the pheromone information. As shown in Figure 1, because the pheromone on the shorter path is accumulated faster than that on the longer path, ants will find a shortest path some time later [15]. In the ACO-based adaptive routing, selection scheme is

---

* Corresponding author. Tel.: +86-02982668672.
E-mail address: meikuizhi@mail.xjtu.edu.cn (K. Mei).

http://dx.doi.org/10.1016/j.micpro.2016.04.001
0141-9331/

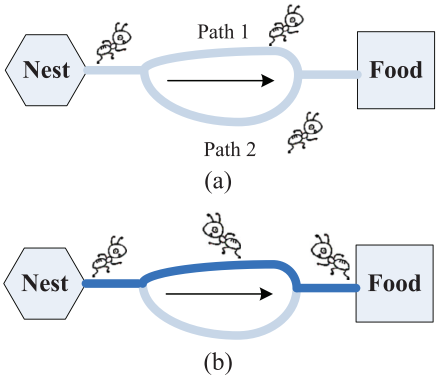
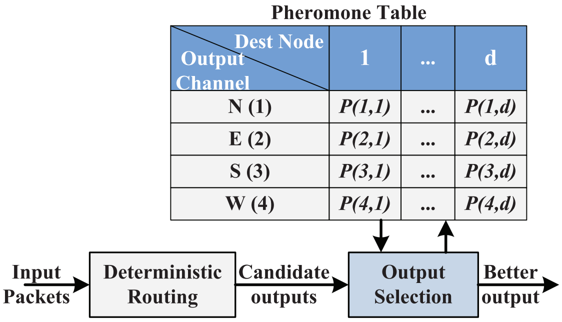


Fig. 1. Diagram of ants' foraging behavior. (a) In the beginning, ants get the food via path 1 and path 2 randomly. (b) Path 1 is shorter than path 2, so more ants bring food back via path 1, and more pheromone is left on the path 1. Finally, path 1 becomes the preferred selection for all ants.



Output Channel \ Dest Node	1	...	d
N (1)	*P(1,1)*	...	*P(1,d)*
E (2)	*P(2,1)*	...	*P(2,d)*
S (3)	*P(3,1)*	...	*P(3,d)*
W (4)	*P(4,1)*	...	*P(4,d)*

Fig. 2. The workflow of a typical ACO-based routing.

implemented by ACO algorithm. With the help of pheromone, it can find a global optimal transmission path for every packet.

For implementing ACO-based routing, each router should have a pheromone table to collect local and global network states. A typical ACO-based routing workflow is shown in Figure 2. Because pheromone is used for selecting output channels, each end-to-end dataflow should be marked by a pheromone, and then the number of pheromones is equal to the number of dataflows. Thus, the size of each pheromone table is about quadratic to the network size. For a large-scale network system, the hardware cost of pheromone tables makes the ACO-based routing algorithms infeasible.

For instance, in a 16 × 16 mesh NoC, there are 1020 ($(16 \times 16 - 1) \times 4$) dataflows from one router to others (data packets transmitted via four directional outputs to 255 destinations), therefore, each pheromone table has 1020 storage cells and each storage cell stores one pheromone value with several bits. In addition to the high memory cost, large pheromone table will cause long table access time and larger power consumption, which have great negative effect on transmission performance.

In this work, we present two techniques for pheromone generating and updating, which can reduce the size of pheromone table with minimum performance penalty:

- Directional Pheromone (DP): In a large-scale network, packet's destination may be far away from the current router, so determining the transmission direction for a packet should consider the traffic load of a region which points to the destination node. Therefore, pheromone can be divided into groups according the directions of destinations.
- Pheromone Updating Sensitivity: In a large-scale network, distant network states are redundant for local pheromone updating and nearby network load prediction, so the two kinds of network states, distant and nearby, should have different impact on pheromone update process [16]. In this paper, we present a variation sensibility parameter, which is changed with each packet's transmission distance, in pheromone feedback update process.

Based on these two features, an ACO-based adaptive routing with directional pheromone (ACO-DP) is proposed for large-scale networks. Our contributions are as follows.

- Combining with the characteristics of travel direction, we propose directional pheromone (DP) for ACO-based routing, which can reduce the size of routing table effectively. Moreover, ACO-DP routing uses a distance-sensitive backward pheromone updating scheme to eliminate the performance degradation caused by the low-cost design.
- A router with ACO-DP routing algorithm is designed and implemented. In our experiments and comparisons, the ACO-DP router achieves highest area efficiency compared with other router designs. For instance, its area efficiency is 2.64 times that of traditional ACO with full size table, and 3.15% more than RACO-CAR when channel queue size is 6 flits. Considering the low-cost routing table and area efficiency, ACO-DP is feasible for large-scale NoC systems.

The remainder of this paper is organized as follows: related works are discussed in Section 2. Then concept of directional pheromone and DP table design are described in Section 3. In Section 4, the design details of ACO-DP-based adaptive routing for mesh NoC systems are given. All experiment results are in Section 5. Section 6 introduces the detailed hardware implementation of an ACO-DP router. Finally, Section 7 concludes this paper.

2. Related works

According to the amount of required network information in output selection, adaptive routing can be divided into three subgroups: regional adaptive, global adaptive, and training-based adaptive routing algorithms.

2.1. Selecting channels with regional information

In consideration of implementation complexity and hardware cost, most adaptive routing algorithms determine transmission path based on regional network load. By checking the size of available flit slots in downstream routers, Output Buffer Length (OBL) selection scheme can choose a reasonable output channel among all adjacent routers [17]. while it makes a routing decision with the current local network information, so the packets cannot avoid network congestion. Similarly, an adaptive deadlock-free routing algorithm, called Dynamic XY (DyXY), has been proposed in [18]. In this scheme, a packet is sent either to the X or Y direction depending on the congestion condition of the neighboring routers. Other region-information-based adaptive routing algorithms include Odd-Even (OE) Model [11], Region-based routing[19], and path-congestion-aware routing [20] and so on.

For detecting the distant congestion region ahead of making an routing decision, a routing with look-ahead congestion detection is presented by using free buffer count at the downstream router [21]. Neighbors-on-Path (NoP) [22] can select next hop with the buffer utilization of neighbor routers. Regional Congestion Awareness (RCA) routing [23] and Destination-Based Adaptive routing [24] get network congestion state by propagating buffer utilization ratio to upstream routers.

The above adaptive routing algorithms determine transmission path at runtime based on the regional network state with low hardware cost. However, distant network states haven't been considered, so the output selections may send too many packets to