

NoC: QoS METRICS MODELLING AND ANALYSIS BASED ON DYNAMIC ROUTING.

Abdelkader SAADAOU¹, Salem NASRI^{1,2}

¹CES-Lab, ENIS, Sfax, Tunisia
gadour3000@hotmail.com

²College of computer, Qassim University, KSA
snasri@qu.edu.sa, salem.nasri@enim.rnu.tn

ABSTRACT

Increasing heterogeneous software and hardware blocks constitute complex ICs known as System on Chip (SoC). These blocks are conceived as intellectual property (IP) cores. Designers are developing SoCs by using IP cores reuse, which include interconnection architecture and interface to peripheral devices. Because of the SoC growing complexity, some researchers tend to concentrate more on the communication rather than the computation aspect. This area of research has leading to the Network on Chip (NoC) Concepts. The research domain of NoC has many applications needing high communication performances. Therefore NoC offers attractive solutions to these applications. One of the goals of NoC technology is to maintain a required Quality of Service (QoS), defined in terms of acceptable parameters values. This paper proposes a presentation of QoS metrics model based on QoS parameters such as End-to-End Delays (EED) and throughputs (Thp), for different applications. This study is based on dynamic routing simulation of a 4x4 mesh NoC behaviour under three communications processes namely TCP, VBR and CBR.

KEYWORDS

NoC, QoS, Dynamic Routing, End-to-End Delay, Throughput, TCP, VBR, CBR.

1. Introduction:

According to ITRS roadmap, the scaling of physical gate length of transistor is reaching less than 10 nm with increasing gap between relative delay of communication (global wire) and computation (gate delay), Figures 1 and 2, [1, 2, 3].

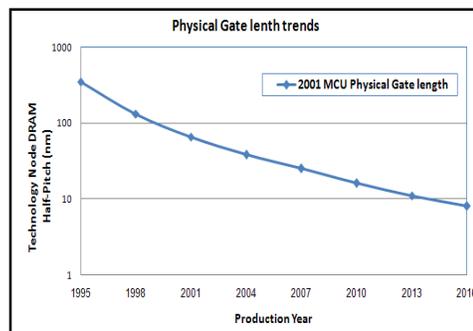


Figure 1: ITRS Roadmap Acceleration Continues-Gate Length Trends

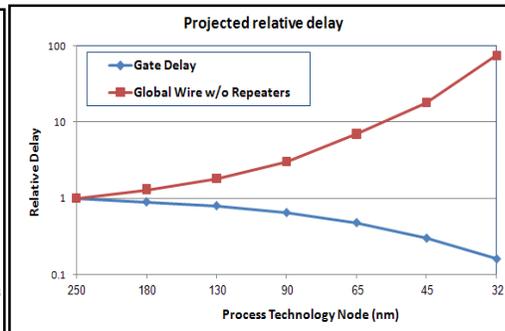


Figure 2: Relative delay for wire and gate vs near future Technologies.

Network on Chip (NoC), as a new SoC paradigm, is useful by handling parallelism, manufacturing complexity, wiring problems and reliability [4, 5, 6]. Researchers have used techniques such as routing and packet-switching concepts of computer networks into a chip [7]. As described in [8, 9, 10, 11], NoCs are emerging as attractive solutions to the existing interconnection constraints implementing future high performance networks and more suitable QoS managements.

Historically, NoC has for origin multi-processors networks [12]. In 1983, transputers permitted the realisation of some parallel machines. In 2000, SPIN (LIP6) constituted the first study of NoC packets commutation using NS2 and systemC tools for simulation. After 2002, researchers have concentrated their effort on bandwidth and latency to guarantee the traffic and to interconnect IPs in the network. Then clock problems were considered. From 2004, many methods and tools of decision (topology, size of FIFO, organization of TDMA) were introduced.

The future moves toward an increasing interaction between operating systems and NoC, with mutual QoS-NoC adaptation of multi-applications.

As a result, many types of NoCs have emerged. Bjerregaard and All describe the most representative NoCs [7].

A NoC is composed by IP cores and routers connected among themselves by communicating channels [8]. Furthermore, packets are composed by header, payload, and trailer. Packets are divided into small pieces called Flits [13, 14]. A flit (Flow control unit) corresponds to the smallest unit of flux control on a link. A phit (Physical unit) corresponds to the quantity of bits that can be transported in one time on the link. The control can be achieved with a granularity of one or several phits.

Nowadays, applications need more performances in direct link with the architecture of the NoC.

This paper presents an overview of this communication centric design paradigm and outlines the scientific efforts made into NoC research area.

Many tentatives to define and modelize QoS metrics were proposed. Bjerregaard and All, were defined QoS as service quantification to the demanding core offered by NoC [7].

Helali and All addressed the problem of metrics for end-to-end QoS management on real time applications by presenting a virtual communication support [15]. Their research was focused on the study of QoS through the switch buffering requirements [16]. In [17] they were interested on NoC switch scheduling and its impact on QoS metrics. Recently Nasri proposed a new approach of QoS metric modelling based on the QoS parameters estimation and applications priority [18, 19].

In this work, we address the QoS metric problem for NoC based system. We propose a new approach of QoS metrics modelling and analysis based on dynamic routing for multi-applications environment with multi parameters.

The next section explains the target NoC architecture. Routing techniques are presented in the section 3, while section 4 focuses on QoS metric modelling requirements. In section 5, we present the experimentation results and analysis. Finally, this paper is ended by conclusions and future works.

2. NoC topology

The topology designates a graph of links between different cores of the NoC [20, 21, 22, 23]. Our choice is built on 4x4 mesh topology as shown in Figure 3.

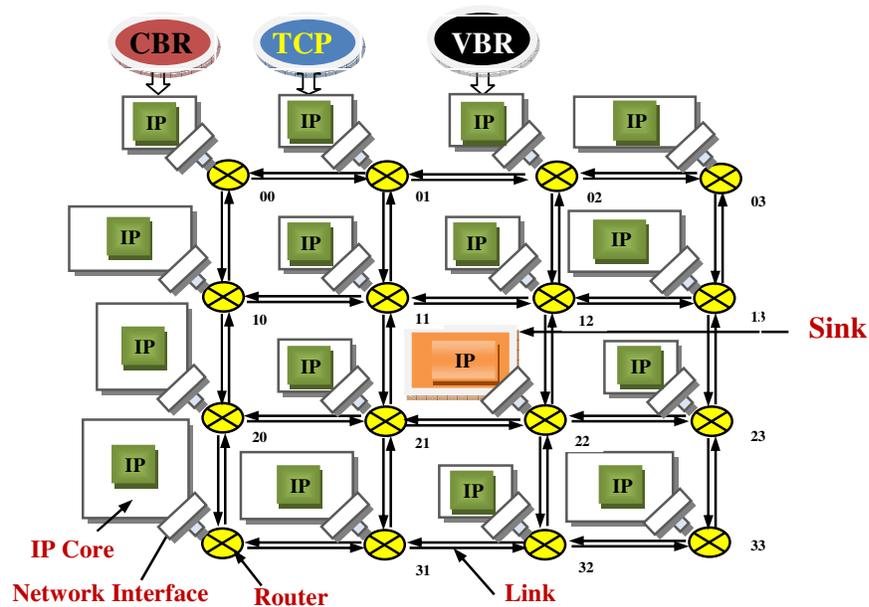


Figure 3: 4x4 Mesh NoC structure.

Each router has five bi-directional ports: East, West, North, South, and Local. The local port used to connect its IP core. The other ports are connected to the neighbor routers. Each router has two (L2), three (L3) or four (L4) bidirectional links with neighbors depending on the position of each one in the graph (Figure 3). In this case study we have considered three different sinks connected to router 33 (L2), router 32 (L3) and router 22 (L4).

3. Dynamic Routing Techniques

The routing algorithms define the path taken by a packet between source and destination [24]. According to where routing decisions are taken, it is possible to classify the routing in source and distributed routing. In source routing, the whole path is decided at the source router, while in distributed routing each router receives a packet and decides about the direction to send it to. According to how a path is defined to transmit packets, routing can be also classified as deterministic or adaptive. In the case of adaptive routing, the path is decided with the progression of the communication [25]. In dynamic routing, the path is a function of the network traffic, which we have used in our simulation [26].

For evaluation of our strategy performance, destinations are linked with three types of routers. We select routers having two (L2), three (L3) or four links (L4) with the same destination.

Trying to meet an ideal network behavior, we define randomly horizontal and vertical failed paths scenarios on the entire NoC, using random broking link duration time. This forces the system to search a new path between sources and destination. Three applications: TCP, VBR, CBR are concurrently active in the same condition and the same time. When horizontal and/or vertical links are tired down, packets go dynamically through other routes taking the shortest path to the destination.

4. QoS METRIC MODELLING REQUIRMENTS

4.1. QoS Definition

Quality of Service (QoS) refers to levels of guarantees given for data transfers. It is a defined measure of performance in a data communications system. For example, to ensure a delivered application such us real-time multimedia without losses information, a traffic contract is negotiated between the network application consumer and provider. This contract guarantees a minimum of bandwidth along with the maximum delay that can be supported.

Since there is no common or formal QoS metrics definition, we propose a new QoS metric approach based on the prioritization factors and parameters. Each application needs different level of performance. Typically QoS parameters include (throughput, end to end delay, jitter, rate of packet loss...). QoS parameters concern also the priority, reliability, speed and amount of traffic sending over a network .

4.2. End to End Delay (EED) and Throughput (Thp)

EED concerns the time for a packet to reach its destination starting from its source. It includes the time elapsed in each node (source- routers) and on links through the communication path until the packet reaches its destination. The delay is in general unpredictable depending on the state of the network. While throughput refers to how much data can be transferred from source to destination in a given amount of time.

4.3. QoS modelling

In a multi-applications environment ($app_1, app_2, \dots, app_m$), we define for each application app_i a set of parameters ($p_{i1}, p_{i2}, p_{i3}, \dots, p_{in}$).

QoS performance parameters should be normalized as \hat{p}_{ij} , with: $p_{ijmax} = \text{Max}\{ p_{ij} \}$ and $p_{ijmin} = \text{Min}\{ p_{ij} \}$, [27].

Then:

- a- For increasing parameters when application value increases:

$$\hat{p}_{ij} = \left| \frac{p_{ij} - p_{ijmin}}{k * p_{ijmax} - p_{ijmin}} \right|$$

- b- For decreasing parameters when application value increases :

$$\hat{p}_{ij} = \left| \frac{p_{ijmax} - p_{ij}}{k * p_{ijmax} - p_{ijmin}} \right|$$

- c- $k \geq 1$: represents the network efficiency coefficient (in our case we chose $k= 1.03$ for example).

If we suppose that we have m applications, QoS can be expressed by the following model:

$$app_1 = \alpha_{11} * \hat{p}_{11} + \alpha_{12} * \hat{p}_{21} + \dots + \alpha_{1n} * \hat{p}_{n1}$$

$$app_2 = \alpha_{21} * \hat{p}_{12} + \alpha_{22} * \hat{p}_{22} + \dots + \alpha_{2n} * \hat{p}_{n2}$$

.....

$$app_m = \alpha_{m1} * \hat{p}_1$$

Then:

$$QoS = QoS_0 + \beta_1 * app_1 + \beta_2 * app_2 + \dots + \beta_m * app_m$$

Hence:

$$QoS = QoS_0 + tr \left(\begin{pmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & \beta_m \end{pmatrix} * \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \vdots & \dots & \ddots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \dots & \alpha_{mn} \end{pmatrix} * \begin{pmatrix} \hat{p}_{11} & \hat{p}_{12} & \dots & \hat{p}_{1m} \\ \hat{p}_{21} & \hat{p}_{22} & \dots & \hat{p}_{2m} \\ \vdots & \dots & \ddots & \vdots \\ \hat{p}_{n1} & \hat{p}_{n2} & \dots & \hat{p}_{nm} \end{pmatrix} \right) \quad (1)$$

Where:

QoS_0 represents the minimum basic required QoS (in our case, we chose $QoS_0 = 10\%$ of the value of the ideal QoS), α_{ij} and β_i are respectively prioritization factors of parameters and applications, arbitrarily fixed referring to the following equations (for one application app_i):

$$\sum_{j=1}^n (\alpha_{ij}) = 1 \quad \text{and} \quad \sum_{j=1}^m (\beta_j) = 1 \quad (2)$$

5. EXPERIMENTATION RESULTS AND ANALYSIS

We used the Ns-2 simulator. It is becoming one of the most popular platforms for performance analysis in the network research community.

In our simulation, we consider different types of router interconnections depending on the position of the router on the NoC. The destination is connected to routers: 22, 32 or 33, which have two (L2), three (L3) or four ports (L4), (Figure 3).

Three applications CBR, VBR and TCP are linked respectively to router 00, 01 and 02. The communication of these applications starts simultaneously in the same time using a dynamic routing. We analyze some QoS metrics such as EED and Thp in the NoC nodes.

5.1. End to End Delay:

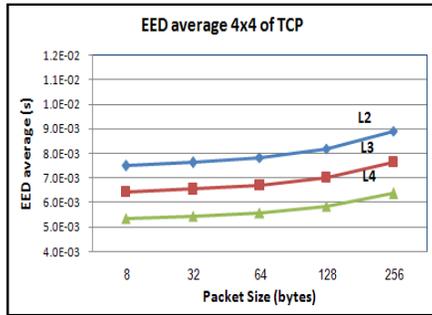


Figure 4: EED average of TCP according to packet size

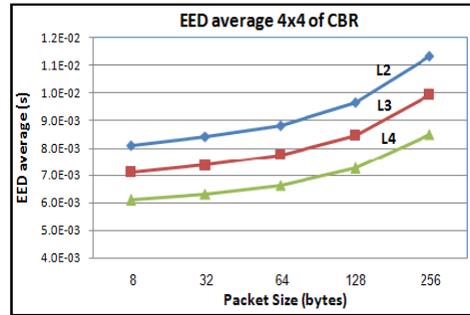


Figure 5: EED average of CBR according to packet size.

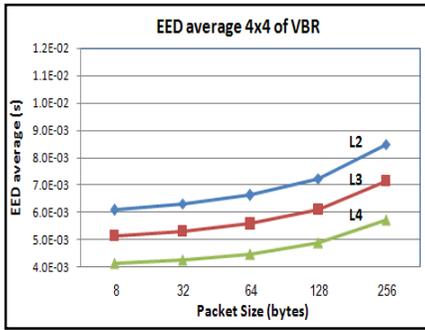


Figure 6: EED average of VBR according to packet size.

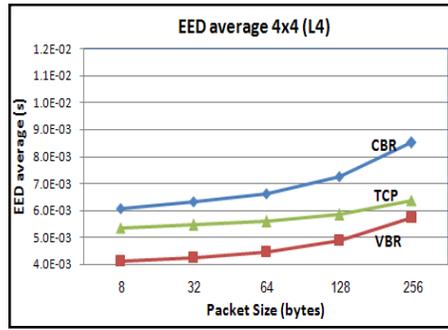


Figure 7: EED average of three Applications according to packet size with router (L4).

Figures 4, 5, 6 and 7 give the relationship between EED average and available packet size. These Figures show that, contrary to CBR, VBR is the application that gives better results. These applications are increasing with EED. Furthermore, the destination linked with a router (L4) presents better result; they also show that application variations are similar in L2, L3 and L4 simulations.

5.2. Throughput (Thp)

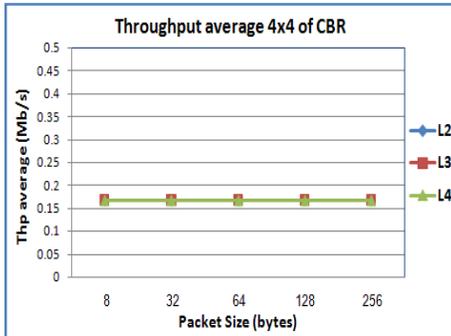


Figure 8: Thp Average of CBR according to packet size.

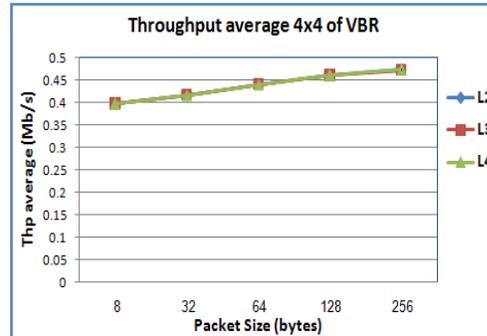


Figure 9: Thp average of VBR according to packet size.

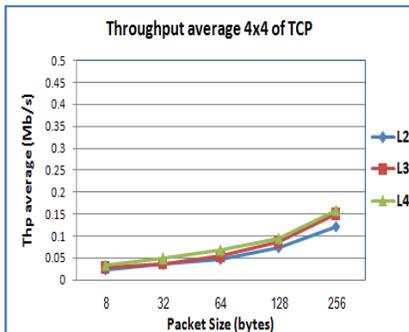


Figure 10: Thp average of TCP applications according to packet size.

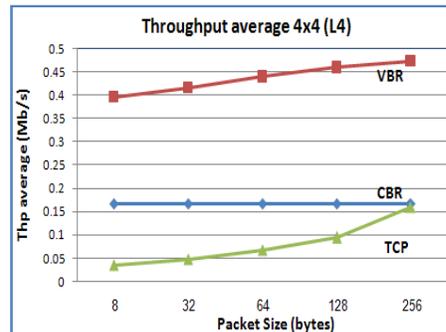


Figure 11: Thp average of three Applications according to packet size with router (L4).

Figures 8, 9, 10 and 11 give the relationship between throughput average and available packet size for CBR, VBR and TCP applications. VBR gives the better throughput average variation than the two other applications. TCP gives worse results. Moreover, CBR is not affected by the

variation of the node type and all applications (TCP, VBR, CBR) are less sensitive to the router position in the NoC.

5.3. QoS measurements and analysis

Referring to the proposed model (1), we choose the parameters α_{ij} , applications β_i prioritization factors and the minimum acceptable value QoS_0 , as shown in Figures 12, 13 and 14. In this model, we consider two QoS performance parameters, EED as p_1 and Thp as p_2 , for three concurrent applications CBR ($i=1$), VBR ($i=2$) and TCP ($i=3$) for different available packet sizes for router type of L4.

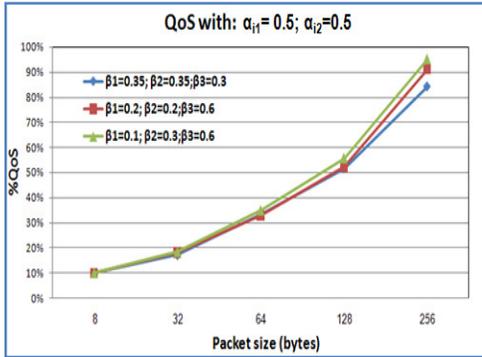


Figure 12: %QoS with parameters

parameters prioritization factors ($\alpha_{i1}=\alpha_{i2}=0.5$) of three applications according to packet size.

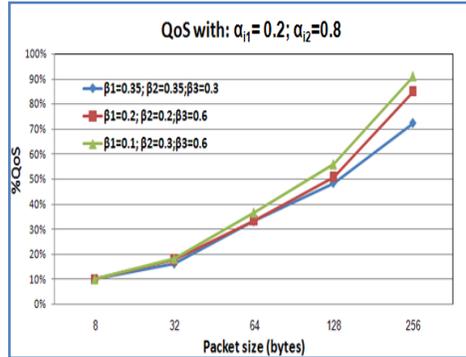


Figure 13: %QoS with

parameters prioritization factors ($\alpha_{i1}=0.2$; $\alpha_{i2}=0.8$) of three applications according to packet size.

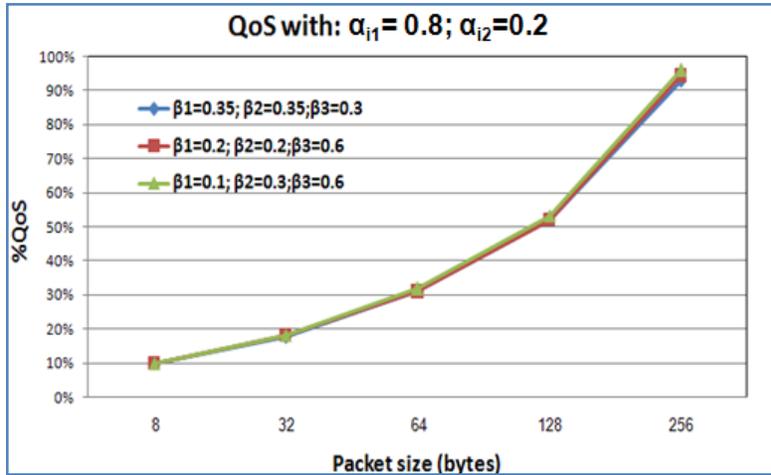


Figure 14: %QoS with parameters prioritization factors ($\alpha_{i1}=0.8$; $\alpha_{i2}=0.2$) of three applications according to packet size.

Figures 12, 13 and 14 show the percent of QoS in relation with the packet size, the scheduling techniques, parameters and applications prioritization factors. It appears that the percent of QoS increases with the packet size. Application prioritization factors have also an impact on the QoS values.

Although the QoS is an abstract notion, we have proposed a new approach of quantifiable representation of the QoS.

6. CONCLUSIONS AND PERSPECTIVES

This paper addresses the QoS metric problem for NoC based system. It proposes a new approach of QoS metrics modelling for network on chip in a multiple applications and multiple parameters environment.

We have focused our study on two fundamental measures of network performances and QoS metrics, EED and Thp in NoC nodes. Two QoS parameters that determine a network connection speed subject of multiple applications in a dynamic routing environment.

Since QoS is qualitative, subjective and not measurable, we have proposed a new approach of its quantifiable representation. In fact, we have shown that metrics of QoS during NoC communication processes are affected by the packet size and its management approach and increased with parameters and applications prioritization factors. This helps to make up the efficiency of the QoS metric evaluation.

QoS metrics measurements based on the router buffer optimization size and load balancing on the NoC with multiple concurrent applications will constitute the future work.

REFERENCES

- [1]International Sematech., (2006) “International Technology Roadmap for Semiconductors”, <http://public.itrs.net>.
- [2]International Sematech., (2003) “International Technology Roadmap for Semiconductors”, <http://public.itrs.net>, interconnect.
- [3]De Micheli G., Seiculescu C., Murali S., Benini L., Angiolini F., Pullini A., (2010) “Networks on chips: From research to products”, in Proc. DAC, pp: 300–305.
- [4]Benini L., De Micheli G., (2002) “ Networks on chips: a new SoC paradigm”, IEEE Computer, V. 35(1), pp: 70-78.
- [5]Al Faruque M. A., Ebi T., Henkel J., (2010) “AdNoC: Runtime Adaptive Network-on-Chip Architecture”, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, , Volume: PP , Issue: 99, pp: 1-1.
- [6]Yiping D., Zhen L., Watanabe T., (2010) “An efficient hardware routing algorithms for NoC”, IEEE, TENCON 2010, pp: 1525 – 1530.
- [7]Bjerregaard T., Mahadevan S., (2006) “A survey of research and Practice of Network on chip”, ACM Computing survey, Vol: 38, Article 1.
- [8]Rijpkema E., Goossens K., Radulescu A., (2003) “ Trade offs in the design of a router with both guaranteed and best-effort services for networks on chip”, Design, Automation and Test in Europe (DATE'03), pp: 350-355.
- [9]Rijpkema E., Goossens K., Wielage P., (2001) “ A router architecture for networks on silicon”, The 2nd Workshop on Embedded Systems (PROGRESS'2001), pp: 181-188.
- [10]Bolotin, E., Cidon, I., Ginosar, R., Kolodny, A., (2004) “QnoC: QoS architecture and design process for network on chip”, Journal of Systems Architecture 50(2-3), Special issue on Networks on Chip, pp:105-128.
- [11]Goossens, K., Dielissen, J., Gangwal, O. P., Gonzalez Pestana, S., Radulescu, A., Rijpkema, (2005) “A design flow for application-specific networks on chip with guaranteed performance to accelerate SoC design and verification”, in DATE, pp:1182-1187.
- [12]Samuel Evain, (2006) “ μ Spider Environnement de Conception de Réseau sur Puce”, thesis, No. D06-16, Institut Nationale des Sciences Appliquées de RENNES (INSA).
- [13]Kim J., Park D., Nicopoulos Ch., Vijaykrishnan N., Das R. Chita., (2005) “Design and analysis of a NoC architecture from performance, reliability and energy perspective”, ACM symposium on

Architecture for networking and communications systems, Princeton, NJ, USA, ISBN:1-59593-082-5, pp:173–182.

[14]Murali S., Coenen M., Radulescu A., Goossens K., De Micheli G., (2006) “A methodology for mapping multiple use-cases onto networks on chips”, Conference on Design, automation and test in Europe, Munich, Germany, ISBN:3-9810801-0-6, pp: 118–123.

[15]A.Helali, A.Soudani, S.Nasri, T.Divaux, (2005) “An approach for end-to-end QoS and network resources management”, Computer Standards & Interfaces Journal vol 28, pp 93-108.

[16]A.Helali I, A.Soudani, S.Nasri, (2006) “Study of Network on chip resources allocation for QoS Management”, Journal of Computer Science vol 2 issue 10, pp 770-774.

[17]Helali A. and Nasri S., (2009) “Network on Chip Switch scheduling Approach for QoS and Hardware Ressources Adaptation”, Int'l J. of Computer Sciences and Engineering Systems (IJCSES), Vol.3, No.1, pp: 29-35.

[18]Nasri S., (2011) “Network on Chip: a New Approach of QoS Metric Modelling Based on Calculus Theory”, Int'l J. of Communications, Network and System Sciences, Vol.3, No.5 ,pp:53-60.

[19]Nasri S., (2011) “New Approach of QoS Metric Modelling on Network on Chip”, Int'l J. of Communications, Network and System Sciences, Vol.4, No.5 ,pp: 351-355.

[20]Jantsch, H. Tenhunen (Eds.), (2003) “Networks on Chip”, Kluwer Academic Publishers, Hingham, MA, USA, ISBN: 978-1-4020-7392-2.

[21]Benini L., De Micheli (Eds.) G., (2006) “Networks on Chips: Technology and Tools”, Morgan Kaufmann Publishers, San Francisco, CA, USA, ISBN: 0123705215.

[22]Dally W.J., Towles B., (2001) “ Route packets not wires: on-chip interconnection networks”, in: Proceedings of Design Automation Conference (DAC'01), ACM, IEEE Press, New York, pp: 648–689.

[23]Winter M., Prusseit S., Gerhard P.F., (2010) “Hierarchical routing architectures in clustered 2D-mesh Networks-on-Chip”, IEEE, SoC Design Conference (ISOCC), pp : 388 – 391.

[24]Yunus S.A.M.J., Marsono M.N., Ibrahim I., (2011) “Modelling router hotspots on Network-on-Chip”, IEEE 13th International Conference on Advanced Communication Technology (ICACT), pp: 896 – 900.

[25]Dally W. J. et Aoki H., (1993) “Deadlock-Free Adaptive Routing in Multicomputer Networks Using Virtual Channels ”, IEEE Trans. Parallel Distrib. Syst., vol. 4, No 4, pp: 466–475.

[26]Baglietto M., Battistelli G., Bolla R., Bruschi R., Davoli F., Zoppoli R., (2005) “ A Neural Network Solution to QoS-IP Team-Optimal Dynamic Routing”, articolo invitato, Proc. Joint 44th IEEE Conference on Decision and Control and European Control Conference, Seville, Spain, pp. 7452-7459.

[27]Yang L., Huacan H., (2007) “ Grid Service Selection Using QoS Model”, Third International Conference on Semantics, Knowledge and Grid (SKG 2007), pp: 576-577.

Authors

Eng. Abdelkader SAADAoui

Received his engineering degree in Electrical Engineering from ‘ENIT: Ecole Nationale d’Ingénieurs de Tunis’ and his Aggregation in Electrical Engineering from the ‘ENSET: Ecole Normale Supérieure de l’Enseignement Technique de Tunis’, Tunisia in 1993 and 1995, respectively. He is simultaneously an assistant professor at the ‘BCT: Buraidah College of Technology’, Kingdom of Saudi Arabia and ‘ISET: Institut Supérieur des études Technologiques de Sfax - Tunisia’. Currently he is preparing his PhD at the ‘ENIS: Ecole Nationale d’Ingénieurs de Sfax - Tunisia’. His research interests include Quality of Service integration in Network on Chip (NoC). His recent work has been in QoS Metrics on multi-



application environment.

Pr. Salem Nasri

Received his PhD in Automatic Control and Computer Engineering from 'INSA: Institut National des Sciences Appliquées' Toulouse, France, in June 1985. He obtained the diploma of "HDR: Habilitation à Diriger les Recherches" in Computer Engineering, in May 2001 from the 'ENIS: Ecole Nationale d'Ingénieurs de Sfax', Tunisia. He is simultaneously Professor at 'ENIM: Ecole Nationale d'Ingénieurs de Monastir', Tunisia, and at the Computer College, Qassim University, Kingdom of Saudi Arabia. His research interests are: Computer Networks, Network on chip, High Speed Protocols, Wireless Communication Systems, Multimedia Applications and Quality of Service Modelling.

