Transport Time Distribution for Deflection Routing on an Odd Torus*

J.M. Fourneau[1] and T. Czachórski[2]

[1] PRiSM, Université de Versailles Saint-Quentin,
45 Av. des Etats Unis, 78000 Versailles, France
[2] IITIS-Polska Akademia Nauk,
Ul. Balticka 5, 44-100 Gliwice, Poland

Abstract. We analyze the performance of all optical packet networks. As optical storage of packets is not available, we assume that the routing protocol is based on deflection. This routing strategy does not allow packets loss. However it keeps the packets inside the network, increases the delay and reduces the bandwidth. Thus the transport delay distribution is the key performance issue for these networks. Here, we consider a 2D torus the size of which is odd. The method is based on a fixed point system between two sub-models. The first subsystem describes the global network performance while the other one models the stochastic behavior of two types of packets.

1 Introduction

All optical packets networks have received considerable attention during the last years. However with actual technology, all-optical networks do not allow the buffering of packets inside the network. Therefore packets have to be sent immediately to the next switch along the path. Old algorithm like Deflection Routing [2] have recently received attention to overcome this weakness [8, 9]. This routing strategy does not allow packets loss. However it keeps the packets inside the network, increase the delay and reduce the bandwidth. In Shortest-Path Deflection Routing, switches attempt to forward packets along a shortest hop path to their destination. Each link can send a finite number of packets per time-slot (the link capacity). If the number of packets which require a link is larger than the capacity, only some of them will use the link they ask for and the other ones have to be misdirected or deflected and they will travel through longer paths. This is the major drawback of this technique.

The tail of the transportation delay and the average usable bandwidth are therefore two major measures of interest. The mean number of deflections is not that large but a significant fraction of the number of packets is heavily deflected. We have observed several packets with more than 1000 deflections during a simulation of a 10×10 2D-mesh with unbalanced traffic [3]. As acknowledgments in

* This work has been partially supported by the French RNRT project *ROMEO*, by a research project between CNRS and Polish Akademia Nauk and by the EURO NGI European Network of Excellence.

J.C. Cunha and P.D. Medeiros (Eds.): Euro-Par 2005, LNCS 3648, pp. 975–983, 2005.

networking protocols must arrive before some timer expiration, heavily deflected packets will be considered as lost because they experience delays larger than the transport time-out. Packets are never physically lost due to physical errors or buffer congestion, but they can be logically lost because the transport delay is too large.

Previous analytical studies of deflection [1, 4] have proposed models for networks based on 2×2 switching blocks without the queueing of new packets. Recently, Fabrega and Muñoz [7] have modeled a network with deflection routing using an approximate model based on Markov chains. However, they have only considered 2×2 switches and a topology such that only one shortest path exists between the source and the destination. Yao et al. have presented in [10] an approximate model for more general topologies which do not contain any directed cycle (again a quite restrictive topology). Clearly, all the models proposed so far have used some unrealistic assumptions about the network topology and switches. Furthermore, all these methods only estimate the mean delay while the important measure is the tail of the delay distribution. Therefore, new methods to obtain the distribution of the delay are still necessary.

In this paper, we consider 4×4 switching elements and a 2D torus topology which was considered as a reasonable topology by the ROM project [8]. We also assume that the size of the torus is odd (2Z +1) and that there are no optical converters. Following the ROM conclusions, we consider fixed size packets and the network is logically synchronous. We model explicitly the routing algorithm with minimal number of deflections per time slot which has been introduced recently by Alcatel [5, 9]. We explain at the end of the paper how to model even size torus.

We model the network by an aggregate representation of the optical packets. First, we represent the vector distance to destination and we gather the packets into two sets according to the number of favorable directions for the next hop. We assume that the packets try to follow a shortest path. Thus only some directions (among the four existing in the torus) are consistent. In an odd torus, we may have packets which have only one possible direction and packets which have two possible directions. In this paper, they are respectively denoted as type 1 and type 2 packets (see Fig. 1). Our analysis is based on the construction of a Markov chain which represents the evolution of a typical packet. The state space takes into account the packet type and the distance vector to the destination.

First, we model the path of a tagged packet inside the net using a Markov chain. The distribution of the transport time can be computed numerically once the deflection probabilities for both types of packets are obtained. The probabilities are the solution of a fixed point system based on the flows of type 1 and type 2 packets. We present an algorithm, numerical results and some simulations to check the accuracy.

This paper is organized as follows: in Section 2, we present the model of a tagged packet based on the topological properties of the torus and the traffic assumptions. Section 3 is devoted to the model of the packet flows. The two sets of equations provide a fixed point system. In Section 4 we present an algorithm and we compare numerical results with simulations.

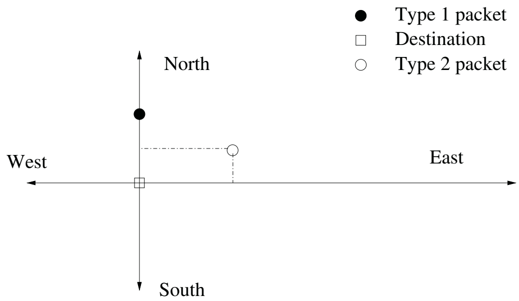

Fig. 1. Type of Packets and Routing

2 Model of a Tagged Packet

Remember that we have divided the packets into two sets: the packets which have only one possible direction (type 1) and the packets which have two possible directions (type 2). A type 1 packet has reached one coordinate of its destination while a type 2 packets must progress in two directions to reach its exit. Of course, at each step, packets may change their types according to their distance to destination and the issue of the deflection algorithm. Borrero and Quessette have proposed the following heuristic which has been proved to be optimal in [5]. The optimality criteria is the number of deflections at each time slot.

Lemma 1 (Degree 1 node of V_1). *[5] If a packet has only one possible direction, we must consider its request before looking at packets of type 2.*

Now, let us turn to the model for routing. We assume that the selection of packet during the routing algorithm is only based on the type of packets. Let $p1$ and $p2$ be respectively the probability that a type 1 and type 2 packet will be deflected at one step. These quantities will be computed in the next section.

Let us now model the evolution of a tagged packet inside the net. First, we build the transition matrix R of a typical packet and we show it for a torus with 7 rows and columns). The model is based on the following set of states: the initial state (state 0) before the packet enters the input node and the completion state (state 1) where the packet leaves the network. State 1 is an absorbing state. We represent explicitly the packet type (denoted as $t1$ and $t2$ in the state description) and the distance vector to destination. Of course, at each step, packets may change their type according to their vector of distances. Due to traffic assumptions, we aggregate states with equivalent vector of distances: for instance $(1, 2)$ and $(2, 1)$. Thus the chain has less states to represent the evolution of the packet inside the net. For instance, for the 7x7 torus, these states are (t1,1,0), (t1,2,0), (t1,3,0), (t2,1,1), (t2,1,2), (t2,1,3), (t2,2,2), (t2,2,3) and (t2,3,3). The chain has 11 states. The transport time is the time of a sample-path beginning at state 0 and finishing at state 1. The PDF of the transport time is obtained by successive multiplications of the distribution probability by transition matrix

R of the Markov chain. The initial distribution puts all the probability mass in state 0.

Now, let us show the effect of the routing algorithm and the deflection on the states. Let us consider the two simplest evolution rules: the non deflected type 1 and the deflected type 2. Remember that the size of the torus is $2Z + 1$.

- A type 1 packet which is non deflected and which is at distance k is kept as a type 1 as it progresses along only one direction. Its distance to destination is therefore $k - 1$.
- A type 2 packet which is deflected remains a type 2 packet. In general, the deflection increases by one the distance to destination, except on the boundary of the torus (see figure 2). Each of the components in the distance vector may be increased with probability 1/2.

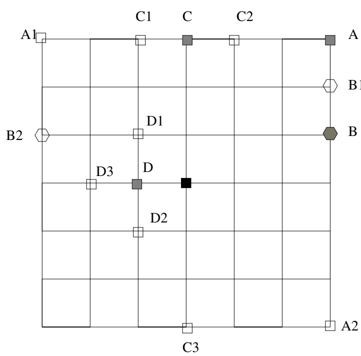

Fig. 2. Deflection on the torus: the destination is the black square. If a packet in A on the boundary is deflected, it will reach nodes $A1$ or $A2$ at next step. But A, $A1$ and $A2$ are at the same distance to destination (i.e. 2Z). For node B, the situation is even more complex, a deflection in B implies that the packet joins node $B1$ or $B2$. At $B2$, the distance is the same, while from node $B1$ it has increased. A deflected type 1 packet in node D may become type 2 (2 possible directions) or stay type 1 (one possible direction). The distance to destination is now increased by one. But if the deflected type 1 comes from the boundary (node C) and it moves to $C3$, then its distance to destination is kept unchanged

Now consider the two other cases: a deflected type 1 packet and a non deflected type 2 packet.

- A type 1 packet at distance k which is deflected has three possible directions. Two directions lead to a type 2 packet (see figure 2) and one direction to a type 1 packet. If we assume equiprobable choices for the directions, the transition rule keeps the packet as type 1 with probability 1/3 and changes it to type 2 with probability 2/3. Its distance to destination is therefore $(1, k)$ except in the following case: when the packet is on the boundary of the torus and it is kept as type 1, the packet is still at distance k along the opposite direction because the torus size is odd.

– A type 2 packet at distance (m, k) which is not deflected decreases its distance to $(m, k-1)$ or $(m-1, k)$. And according to its position and the direction selected it may become a type 1 (if $m-1=0$ or $k-1=0$) or stay a type 2 packet otherwise.

Thus the transitions in R can be easily obtained from the rules formerly shown and the deflection probabilities $p1$ and $p2$. The last part of the matrix still missing gathers the transitions out of states 0 (the arrivals inside the net). Each destination node in the net (except the source of the packet) has the same probability (i.e. $\frac{1}{N^2-1}$) to be the destination. So, we must count the number of nodes of type 1 and 2 at distance $(k, 0)$ or (m, k).

$$R(0,(t1,k)) = \frac{4}{N^2-1} \quad and \quad R(0,(t2,m,k)) = \frac{x2(k+m)}{N^2-1} \tag{1}$$

where $x2(k+m)$ is the number of nodes in the torus at distance k of their destination which may contain packets of type 2. Clearly, $x2(m+k) = 4k+4m-4$ if $k+m \leq Z$ and $x2(m+k) = 8Z+4-4k-4m$ when $Z \leq m+k \leq 2Z$. Finally for the 7×7 torus we get (with $q1 = 1-p1$ and $q2 = 1-p2$):

$$R = \left[\begin{array}{c|ccc|cccccc}
 & 1/12 & 1/12 & 1/12 & 1/12 & 1/6 & 1/6 & 1/12 & 1/6 & 1/12 \\
1 & & & & & & & & & \\
\hline
q1 & & p1/3 & & 2p1/3 & & & & & \\
 & q1 & & p1/3 & & 2p1/3 & & & & \\
 & & q1 & p1/3 & & & 2p1/3 & & & \\
\hline
 & q2 & & & & p2 & & & & \\
 & & q2/2 & & q2/2 & & p2/2 & p2/2 & & \\
 & & & q2/2 & & q2/2 & p2/2 & & p2/2 & \\
 & & & & & q2 & & & p2 & \\
 & & & & & & q2/2 & q2/2 & p2/2 & p2/2 \\
 & & & & & & & & q22 & p2
\end{array}\right]$$

3 Macroscopic Model of the Flows

Clearly, the first model does not take into account the arrival process because we assume that the packet is in the network. We now study the flow of packets. Note that due to the topology and the traffic assumptions all the switches are statistically equivalent. The probabilities of deflection are computed by conditioning on the arrivals. Then these probabilities are shown to be related to the load of the link using an independence assumption. These last relations provide a fixed point system for $p1$ and $p2$.

3.1 Deflection Probability

The best routing algorithm must route type 1 packets with a higher priority [9]. Therefore it is sufficient to compute the deflection probability of a tagged type 1 packet knowing the exact number of type 1 packets in the switch. Note that the upper bound of the index is 3 because the tagged customer uses one input link of the switch.

$$p1 = \sum_{i=0}^{3} Pr(i\ type\ 1\ arrivals) d1(i)$$

where $d1(i)$ is the probability that the tagged packet of type 1 will be deflected if another type 1 packet arrives. The probabilities of arrivals are obtained by an independence assumption. Let us denote by u_1 the utilization of an arbitrary link by type 1 packets.

$$Pr(i\ type\ 1\ arrivals) = C(3,i)(u_1)^i(1-u_1)^{3-i} \tag{2}$$

As type 2 customers have a lower priority in the routing algorithm, their deflection probability requires a conditioning on a more complex set of arrivals.

$$p2 = \sum_{i=0,j=0}^{i+j=3} Pr(i\ type\ 1\ and\ j\ type\ 2\ arrivals) d2(i,j) \tag{3}$$

where $d2(i,j)$ denotes the probability that the tagged type 2 packets will be deflected due to the arrivals of i type 1 and j type 2 packets. Similarly the probability of arrivals follows a multinomial distribution because of the independence assumption ($B(3,i,j)$ is the multinomial coefficient):

$$Pr(i\ type\ 1\ and\ j\ type\ 2\ arrivals) = B(3,i,j)(u_1)^i(u_2)^j(1-u_1-u_2)^{3-i-j}$$

We now have to obtain the elementary probabilities $d1(i)$ and $d2(i,j)$. First we consider an arbitrary tagged type 1 packet entering into an arbitrary switch. Clearly, $d1(0) = 0$ and $d2(0,0) = 0$ because there is no competition and $d2(1,0) = 0$, and $d2(0,1) = 0$ as a type 1 packet or a type 2 packet is not sufficient to deflect another type 2 packet. To compute the other values, we assume equiprobable choices when several packets of the same type request the same output. For the sake of conciseness, we omit the computation of $d1$ and $d2$ (see [6] for more details) and we give the results for positive values in Table 1. During this computation, we take care of some properties of routing on an odd torus. For instance, we have:

Lemma 2. *All configurations of requests are not possible due to the routing algorithm and the topology. For shortest path deflection routing in an odd torus, a type 2 packet can not ask for two opposite directions (for instance North and South).*

3.2 Average Distance and Deflection Probabilities

Let us now establish new relations between the link utilization u and the deflection probabilities. We must consider now the number of packets $\vec{n1}(k)$ and

Table 1. Table for $d1$ and $d2(i,j)$

d1(1)	d1(2)	d1(3)	d2(0,2)	d2(0,3)	d2(1,1)	d2(1,2)	d2(2,0)	d2(2,1)	d2(3,0)
$\frac{1}{8}$	$\frac{11}{48}$	$\frac{81}{256}$	$\frac{1}{8}$	$\frac{9}{32}$	$\frac{1}{16}$	$\frac{15}{64}$	$\frac{1}{48}$	$\frac{23}{192}$	$\frac{13}{128}$

$\vec{n2}(m,k)$ rather than the state of a single tagged packet. However, the evolution is modeled by a stochastic matrix M that we can deduce from R. For transitions inside the network, $M(i,j)$ is equal to $R(i,j)$. Indeed, the average numbers of customers obey the same evolution rules than a single packet. The only differences are in the transition between the network and the outside which reflects the arrival rate. We remove the first two states from R and modify the first row to take into account the flow entering the network. We assume Poisson arrivals with rate λ. We need to compute the number of fresh packets of type 1 or 2 entering the network at distance $(k,0)$ or (m,k). Let us denote $a1(k)$ and $a2(m,k)$ these numbers. The average number of packets entering the network is also the average number of packets entering the electronic buffers, if the system is stable. Therefore it is equal to λN^2.

$$a1(k) = \lambda N^2 R(0,(t1,k)) \quad and \quad a2(m,k) = \lambda N^2 R(0,(t2,m,k)) \tag{4}$$

But, the flow entering the network must be equal to the flow leaving the switches with a successful transition from a node at distance 1. Therefore: $\lambda N^2 = \vec{n1}(1) * (1-p1)$. Finally, the average population vector is the solution of the linear system:

$$(\vec{n1},\vec{n2}) = (\vec{n1},\vec{n2})M \quad and \quad \vec{n1}(1) = \frac{\lambda N^2}{(1-p1)} \tag{5}$$

Let us now turn back to the link utilization $u1$ and $u2$. As all the links are equivalent due to the topology and the traffic assumptions, we get:

$$u1 = \frac{\sum_{k=1}^{Z} \vec{n1}(k)}{4N^2} \quad and \ u2 = \frac{\sum_{k=2}^{2Z} \vec{n2}(k)}{4N^2}$$

Thus we obtain a fixed point. We have proved the existence of a solution to this system using Brouwer's fixed point theory, the continuity of steady-state distribution proved by Malyšev and the convexity of $p1$ and $p2$ (see equations 2 and 3). For the sake of conciseness, the proof is omitted (see [6] for more details).

4 Experimental Results

Let us now turn to the numerical algorithm. The computation is iterative: at each step we compute the new transition probability matrix M and a new set of values for $\vec{n1}(i)$ and $\vec{n2}(i,j)$. Then we get $p1$ and $p2$ and compute the difference with their former values. If this difference is smaller than 10^{-9}, we stop the iterations. Initially, $p1$ and $p2$ are equal to 0. The computation of vectors $\vec{n1}$ and $\vec{n2}$

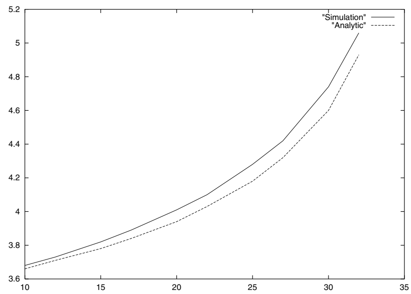

Fig. 3. Comparison of simulation and analytical results: average transport time (in hops) versus load (in packets arriving in the global networks per time slot)

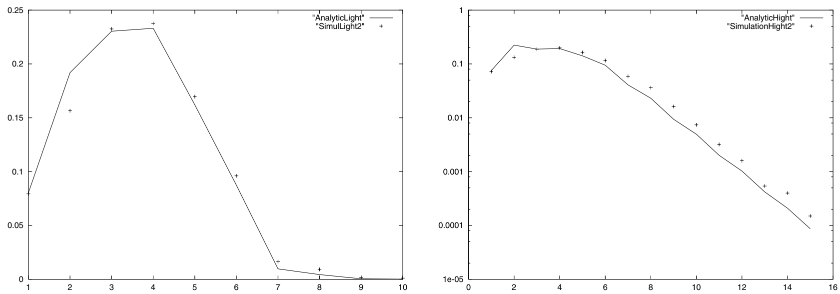

Fig. 4. Comparison of simulation and analytical results: distribution of the transport time for arrival rate=10 (left) and 25 (right)

is performed by an usual direct elimination solver. The number of iterations is usually very low and the matrix is very small: we observe a convergence before ten iterations. Once the fixed point is found, we obtain the transport time distribution using the probabilities $p1$ and $p2$ and few vector matrix multiplications.

We compare the numerical results obtained by our approach to simulation results for a 7×7 torus. In Fig. 3, we present the evolution of the average transport time (in hops) versus the global arrival rate (in number of packets for the whole network). Clearly, the approximations are quite accurate, even if we have used a large scale to emphasis the difference. The analytical results look optimistic. We now present the distribution of the transport time at light and moderate load (Fig. 4). Again we depict the simulation and the analytical results. And the figures show the accuracy of our method, even for the distribution of the transport delay.

5 Conclusions

To the best of our knowledge, it is the first approach to analyze the transport time distribution for more general switches and torus. It is possible to model even torus instead of odd ones. We must slightly change the first model. In an even torus, packets may have more than 2 good directions and the probabilities used to define matrix R are slightly different. To complete the approach, one must also study the distribution of the waiting time before entering the network. Diffusion models of these queues are currently under development.

References

1. J. Bannister, F. Borgonovo, L. Fratta, M. Gerla, A versatile model for predicting the performance of deflection routing networks, Performance Evaluation, V16, pp 201-222, 1992.
2. P. Baran. On distributed communication networks. *IEEE Transactions on Communication Systems*, CS-12, 1964.
3. D. Barth, P. Berthomé, A. Borrero, J.M. Fourneau, C. Laforest, F. Quessette, and S. Vial. Performance comparisons of Eulerian routing and deflection routing in a 2d-mesh all optical network. In *ESM'2001*, 2001.
4. A. Bonnoni, P.P. Prucnal, Analytical evaluation of improved access techniques in deflection routing networks, IEEE/ACM Trans. on Networking, V4, N5, 1996.
5. A. Borrero, J.M. Fourneau and F. Quessette Packet Selection in a deflection routing algorithm. *ISCIS 2002*, Orlando, USA, CRC Press.
6. T. Czachórski and J.M. Fourneau, Performance evaluation of an optimal deflection routing algorithm on an odd torus, Archiwum informatyki teoretycznej i stosowanej, V 16, N 4, pp 257-277, 2004.
7. J. Fabrega and X. Muñoz, A study of Network Capacity under Deflection routing schemes, Europar 03, LNCS 2790, pp 989–994.
8. P. Gravey, S. Gosselin, C. Guillemot, D. Chiaroni, N. Le Sauze, A. Jourdan, E. Dotaro, D. Barth, P. Berthomé, C. Laforest, S. Vial, T. Atmaca, G. Hébuterne, H. El Biaze, R. Laalaoua, E. Gangloff, and I. Kotuliak. Multiservice optical network: Main concepts and first achievements of the ROM program. *Journal of Ligthwave Technology*, 19:23–31, January 2001.
9. S. Mneimeh and F. Quessette. *Minimum Deflection Routing Algorithm*, Alcatel Patent Application #135945, 2002
10. S. Yao, B. Mukherjee, and S. Dixit. *Plato: a generic modeling technique for optical packet switched networks* International Journal on Wireless and Optical Communications, V1, N1, 2003, pp 91-101. World Scientific Publishing.