Deflection Routing on a Torus is Monotone *

Ana Bušić, Jean-Michel Fourneau, and David Nott

PRiSM, Université de Versailles Saint-Quentin-en-Yvelines,
45 Av. des Etats Unis, 78000 Versailles, France,
Email: {abusic, jmf, dnott}@prism.uvsq.fr

Summary. Deflection Routing is proposed for all optical core switches because of the lack of optical memory. In Shortest-path Deflection Routing, switches attempt to forward packets along a shortest hop path to their destinations. Each link can send a finite number of packets per time-slot (the link capacity). Incoming packets have to be sent immediately to their next switch along the path. If the number of packets which require a link is larger than the link capacity, only some of them will use the link they ask for and the other ones have to be misdirected or deflected. We build the Markov chain which models a packet routing in an odd torus. We prove that this matrix is $\preceq_{st}$-monotone. The proof is based on increasing sets as we consider a partial ordering on the state space based on the network topology.

1 Introduction

All optical packet networks have received considerable attention during the last years due to the high bandwidth they could offer. But the lack of optical memory prohibits the buffering of packets inside the network and the use of "store and forward" routing algorithms. Deflection Routing [1] and Convergence Routing [4] have been developed to overcome this weakness. These routing strategies do not lose packets but they keep them 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 destinations. Each link can send a finite number of packets per time-slot (the link capacity). Incoming packets have to be sent immediately to their next switch along the path. If the number of packets which require a link is larger than the link 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 on longer paths. Packets will eventually never reach their destination (i.e. the livelock problem). The tail of the transportation delay is therefore a major measure of

* This work has been partially supported by the French RNRT project *ROMEO* and the French "programme blanc" project *SMS*.

interest. Previous analytical studies of deflection algorithms and networks are based on approximate Markovian model of very simple topologies and switching elements (see [3] for a review). Recently a fixed point system based on two Markov models has been proposed [3]. One of these models describes the end to end delay of a tag packet when the deflection probabilities are known. The other model allows the computation of the deflection probabilities when the link capacity is 1.

Here we prove that the Markov chain associated to the end to end delay is monotone. Thus we can find bounds on the end to end delay when we replace an unknown value of the deflection probability by an upper bound. First in Sect. 2 we briefly introduce stochastic ordering on a partially ordered space. Then we present in Sect. 3 the routing, the topology and the proof of the main theorem on stochastic monotonicity. Finally in Sect. 4 we show using the monotonicity theorem that the end to end delay is stochastically increasing and that the necessary condition on the traffic to obtain the monotone property is satisfied for saturation traffic under some technical constraint. We must emphasize here that the main result allows to bound the delay for any traffic using the monotone matrix as an upper bound. We can use our result to guarantee performance for larger link capacity and more complex traffic.

2 Stochastic ordering on a partially ordered space

We will first give the definition of $\preceq_{st}$-comparison of two random variables defined on a partially ordered space. Then we will introduce the monotonicity property for a transition matrix of a homogeneous discrete time Markov chain (DTMC).

Recall that a binary relation $\preceq$ on a set $\mathcal{S}$ is called a partial order on $\mathcal{S}$ if it reflexive, transitive, and antisymmetric. If additionally for all $x, y \in \mathcal{S}$ either $x \preceq y$ or $y \preceq x$ holds, then the relation $\preceq$ is called a total order on $\mathcal{S}$. A typical example of a partial order that is not a total order is the product order on a product space. Let $\mathcal{S}$ be $\mathbb{N}^I$ or $\mathbb{R}^I$, where I is a countable set. The product order $\preceq$ on $\mathcal{S}$ is defined by $x, y \in \mathcal{S},\ x \preceq y$ if $x_i \leq y_i, \forall i \in I$.

Definition 1. *Let $\mathcal{S}$ be a space endowed with a partial order $\preceq$ and let X and Y be two random variables on $\mathcal{S}$. X is smaller than Y in a strong stochastic sense, $X \preceq_{st} Y$, if*

$$\mathbf{E}[f(X)] \leq \mathbf{E}[f(Y)], \text{ for each increasing function } f,$$

provided that the expectations exist.

A subset $U \in \mathcal{S}$ is called an increasing (upper) set if its indicator function $\mathbf{1}_U$ is increasing. It follows that U is an increasing set if and only if $x \in U$ and $x \preceq y$ imply $y \in U$. The following characterization (see [6] for a proof) is often used as definition of $\preceq_{st}$-order on a partially ordered space $\mathcal{S}$.

Proposition 1. *$X \preceq_{st} Y$ if and only if $P(X \in U) \leq P(Y \in U)$, for all increasing sets $U \subset \mathcal{S}$.*

In the following we consider some properties of transition matrices of homogeneous DTMCs on a finite partially ordered state space $(\mathcal{S}, \preceq)$.

Definition 2. *A transition matrix P is monotone if for all probability vectors u and v, $u \preceq_{st} v$ implies $uP \preceq_{st} vP$.*

Let us denote by $P_{x,*}$ row x of transition matrix P. Then the $\preceq_{st}$-monotonicity can be characterized as follows (see [6]).

Proposition 2. *A transition matrix P is $\preceq_{st}$-monotone if for all $x, y \in \mathcal{S}$ such that $x \preceq y$, $P_{x,*} \preceq_{st} P_{y,*}$, i.e. if $\sum_{k \in U} P_{x,k} \leq \sum_{k \in U} P_{y,k}$ for all increasing sets U.*

Let P and Q be transition matrices of two homogeneous DTMCs on the same state space.

Definition 3. *We say that $P \preceq_{st} Q$ if $P_{x,*} \preceq_{st} Q_{x,*}$ for all $x \in \mathcal{S}$, i.e. if $\sum_{k \in U} P_{x,k} \leq \sum_{k \in U} Q_{x,k}$ for all increasing sets U.*

The monotonicity property and stochastic comparison of transition matrices give sufficient conditions for comparison of two DTMCs.

Theorem 1. *(see [6] for a proof) Let $(\mathcal{S}, \preceq)$ be a partially ordered space and let $\{X_n\}_{n \geq 0}$, $\{Y_n\}_{n \geq 0}$ be two DTMCs with transition matrices P and Q. If $X_0 \preceq_{st} Y_0$, at least one transition matrix P or Q is $\preceq_{st}$-monotone, and $P \preceq_{st} Q$, then $\{X_n\}_{n \geq 0} \preceq_{st} \{Y_n\}_{n \geq 0}$, i.e. $X_n \preceq_{st} Y_n,\ for\ all\ n \geq 0$.*

3 Deflection Routing is $\preceq_{st}$-monotone

Consider a 2D odd torus (see Fig. 1) of size $N = 2M + 1$ with uniform traffic. We are interested in end to end delay of an arbitrary tag packet. It has been proved in [5] that the more efficient routing algorithms route with a higher priority the packets which have only one good direction to follow on the torus. Remember that we assume that the packets try to follow a shortest path. In an odd torus and a grid, packets may have one or two possible directions. In this paper, packets are respectively denoted as type 1 and type 2 packets. A type 1 packet has reached one coordinate of its destination while a type 2 packet 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. Their deflection probabilities are denoted as p_1 and p_2. We model the optimal routing algorithm described in [2].

The state of the tag packet is its distance vector to destination. As with packets, states may be of type 1 or 2. Due to the symmetry of torus

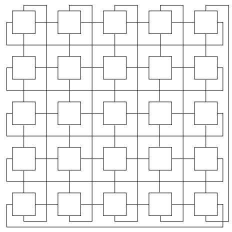

Fig. 1. A 2D torus of size 5

and traffic assumptions, we can aggregate states with equivalent vector of distances: for instance $(1,2)$ and $(2,1)$. The state space of the model is $\mathcal{S} = \left\{x = (x_1, x_2) \mid 0 \leq x_2 \leq x_1 \leq M\right\}$, where $M = \left\lfloor \frac{N}{2} \right\rfloor$. We will denote by $\mathcal{S}^* = \mathcal{S} - \{(0,0)\}$.

Following [3] we can build the transition matrix using four rules to describe the evolution of both types of packet when they are deflected or when they succeed. Remember that the size is odd and take care of the boundary of the torus. The behaviour of the boundary cases may be found in [3].

- A type 1 packet which is not 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.
- A type 1 packet at distance k which is deflected has three possible directions. One direction leads to a type 1 packet (at distance k+1) and two directions lead to a type 2 packet (at distance (k,1)).
- 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.

We obtain the following matrix for a 7×7 torus (with $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$). The states are ordered according to their type (1 or 2) and then according to the distance to destination.

$$
\left[\begin{array}{c|ccc|cccccc}
1 & & & & & & & & & \\
\hline
q_1 & & p_1/3 & & 2p_1/3 & & & & & \\
 & q_1 & & p_1/3 & & 2p_1/3 & & & & \\
 & & q_1 & p_1/3 & & & 2p_1/3 & & & \\
\hline
 & q_2 & & & & p_2 & & & & \\
 & & q/2 & & q_2/2 & & p_2/2 & p_2/2 & & \\
 & & & q_2/2 & & q_2/2 & p_2/2 & & p_2/2 & \\
 & & & & & q_2 & & & p_2 & \\
 & & & & & & q_2/2 & q_2/2 & p_2/2 & p_2/2 \\
 & & & & & & & & q_2 & p_2
\end{array}\right]
$$

Note that this matrix is not $\preceq_{st}$-monotone under this total order on the state space. We will consider the classical product partial order $\preceq_\star$ on $\mathcal{S}$:

for $x = (x_1, x_2)$, $y = (y_1, y_2) \in \mathcal{S}$,

$$x \preceq_\star y \text{ if } x_i \le y_i,\ i = 1, 2.$$

We use an event based description of the chain. Suppose that the actual position of the packet is $x \in \mathcal{S}^*$. We will denote by d the direction that corresponds to the first component. We distinguish the following events:

1. The packet is deflected in the direction $\bar{d}$ that is opposite to d (event E_1).
2. The packet is deflected in other direction than $\bar{d}$ (event E_2).
3. The packet is not deflected and it takes direction d (event E_3).
4. Only for type 2 states: the packet is not deflected and it takes the other direction than d (event E_4).

Remark that, for type 1 states, event E_2 is in fact the aggregated version of events corresponding to two different remaining deflection directions. However, due to the symmetry of torus, both events induce the same transition $(x_1, 0) \to (x_1, 1)$. The probabilities of events are different for type 1 ($x_2 = 0$) and type 2 states ($x_2 > 0$) and are given in Table 1.

Type of state	$P(E1)$	$P(E2)$	$P(E3)$	$P(E4)$
Type 1 ($x_2 = 0$)	$\frac{p_1}{3}$	$\frac{2p_1}{3}$	q_1	$-$
Type 2 ($x_2 > 0$)	$\frac{p_2}{2}$	$\frac{p_2}{2}$	$\frac{q_2}{2}$	$\frac{q_2}{2}$

Table 1. Probabilities of events for type 1 and type 2 states.

It remains us to study the transitions induced by each event. Events E_1 and E_4 can be simply described by functions $f_1 : \mathcal{S}^* \to \mathcal{S}$ and $f_4 : \mathcal{S} - \{x \mid x_2 = 0\} \to \mathcal{S}$,

$$f_1(x_1, x_2) = \big(\min\{M, x_1 + 1\}, x_2\big),$$
$$f_4(x_1, x_2) = (x_1, x_2 - 1).$$

Remark that $f_1(M, x_2) = (M, x_2)$ is a consequence of the structure of an odd torus. For events E_2 and E_3 we need to consider separately the case when $x_1 = x_2 = t$ because of aggregation of positions with distance vectors $(t, t+1)$ and $(t+1, t)$ into one state $(t+1, t)$. Thus, we can describe E_2 and E_3 by functions $f_2, f_3 : \mathcal{S}^* \to \mathcal{S}$,

$$\begin{aligned} f_2(x_1, x_2) &= \begin{cases} (x_1, \min\{M, x_2 + 1\}), & x_1 > x_2 \\ (\min\{M, t+1\}, t), & x_1 = x_2 = t \end{cases} \\ &= \big(\max\{x_1, \min\{M, x_2 + 1\}\},\ \min\{x_1, \min\{M, x_2 + 1\}\}\big), \\ f_3(x_1, x_2) &= \begin{cases} (x_1 - 1, x_2), & x_1 > x_2 \\ (t, t-1), & x_1 = x_2 = t \end{cases} \\ &= \big(\max\{x_1 - 1, x_2\},\ \min\{x_1 - 1, x_2\}\big). \end{aligned}$$

The following lemma follows easily from the fact that the operators min and max are order preserving.

Lemma 1. *Functions f_1, f_2, f_3, and f_4 are increasing functions under the product partial order $\preceq_\star$.*

Let us remark that functions f_i correspond to events $E_i, i = 1 \ldots 4$ and not directly to different transitions. Indeed, in the case $x_1 = x_2$, we have only two possible transitions as $f_1(x) = f_2(x) = (\min\{M, x_1 + 1\}, x_1)$ and $f_3(x) = f_4(x) = (x_1, x_1 - 1)$. We will now show that the transition matrix of considered model is $\preceq_{st}$-monotone with respect to the partial order $\preceq_\star$ and under the hypothesis that deflection probabilities p_1 and p_2 verify $p_1 \leq p_2$. This hypothesis is needed to assure the comparison of rows of the matrix corresponding to two states $x, y \in \mathcal{S}^*$ such that $x \preceq_\star y$, $x_2 = 0$ (type 1 state) and $y_2 > 0$ (type 2 state).

Theorem 2. *Assume that the deflection probabilities p_1 and p_2 verify $p_1 \leq p_2$. Then the transition matrix of the end to end delay for a Deflection Routing in an odd torus is $\preceq_{st}$-monotone with respect to the partial order $\preceq_\star$.*

Proof. Let x and y be two arbitrary states from $\mathcal{S}$ such that $x \preceq_\star y$. We need to show that $P_{x,*} \preceq_{st} P_{y,*}$ (see Proposition 2). After Proposition 1 this is equivalent to show that $\sum_{k \in U} P_{x,k} \leq \sum_{k \in U} P_{y,k}$ for all increasing sets U. Let us remark first that $P_{(0,0),*} \preceq_{st} P_{x,*}, \forall x \in \mathcal{S}$ since $(0,0)$ is an absorbing state and $(0,0) \preceq_\star x, \forall x \in \mathcal{S}$. Suppose now that $x \neq (0,0)$. We distinguish the following cases:

1. both x and y are type 1 states, i.e. $x_2 = y_2 = 0$,
2. both x and y are type 2 states, i.e. $0 < x_2 \leq y_2$,
3. x is type 1 and y type 2 state, i.e. $0 = x_2 < y_2$.

The proof for the first two cases follows directly from Lemma 1 and the fact that, for the two states of the same type, the probability of each event is constant (Table 1). Indeed, if for $i \in \{1, \ldots, 4\}$, $f_i(x) \in U$ for an increasing set U, then $f_i(y) \geq f_i(x)$ since f_i is increasing, so $f_i(y) \in U$ by definition of an increasing set. Thus, $\sum_{k \in U} P_{x,k} \leq \sum_{k \in U} P_{y,k}$ for all increasing sets U.

Let us consider now the third case. We have: $f_1(x) = (\min\{M, x_1 + 1\}, 0)$, $f_2(x) = (x_1, 1)$, and $f_3(x) = (x_1 - 1, 0)$. We can notice that

$$f_3(x) \preceq_\star f_1(x) \text{ and } f_3(x) \preceq_\star f_2(x). \tag{1}$$

On the other hand, from Lemma 1 we have

$$f_i(x) \preceq_\star f_i(y), \forall i \leq 3. \tag{2}$$

Additionally,

$$f_2(x) \preceq_\star f_1(y),\ f_3(x) \preceq_\star f_4(y). \tag{3}$$

We have three different types of increasing sets U for which $\sum_{k \in U} P_{x,k} > 0$:

- $f_3(x) \in U$. Then (1) implies $f_1(x), f_2(x) \in U$. On the other hand, (2) and (3) give $f_i(y) \in U, \forall i$, so $\sum_{k \in U} P_{x,k} = \sum_{k \in U} P_{y,k} = 1$.

- $f_3(x) \notin U,\ f_2(x) \in U$. Then (2) and (3) imply $f_2(y),\ f_1(y) \in U$. Thus by hypothesis $p_1 \leq p_2$, $\sum_{k \in U} P_{x,k} \leq p_1 \leq p_2 \leq \sum_{k \in U} P_{y,k}$.
- $f_2(x), f_3(x) \notin U,\ f_1(x) \in U$. Then (2) and hypothesis $p_1 \leq p_2$ imply $\sum_{k \in U} P_{x,k} = \frac{p_1}{3} \leq \frac{p_2}{2} \leq \sum_{k \in U} P_{y,k}$. Thus, $P_{x,*} \preceq_{st} P_{y,*}$. □

4 End to end delay

We will show that end to end delay is stochastically increasing in deflection probability parameter vector $p = (p_1, p_2)$.

Theorem 3. *Let $p = (p_1, p_2)$ and $r = (r_1, r_2)$ be two deflection probability parameter vectors such that $p_1 \leq r_1$ and $p_2 \leq r_2$, and let $\{X_n\}_{n \geq 0}$ and $\{Y_n\}_{n \geq 0}$ be respectively the two DTMCs describing Deflection Routing on a torus with parameter vectors p and r. If $X_0 \preceq_{st} Y_0$, and at least one of $p_1 \leq p_2$ or $r_1 \leq r_2$ holds, then $\{X_n\}_{n \geq 0} \preceq_{st} \{Y_n\}_{n \geq 0}$.*

Proof. Let P and Q denote respectively transition matrices of chains $\{X_n\}_{n \geq 0}$ and $\{Y_n\}_{n \geq 0}$. Theorem 2 and $p_1 \leq p_2$ or $r_1 \leq r_2$ imply that at least one transition matrix, P or Q, is $\preceq_{st}$-monotone. It remains us to show that $P \preceq_{st} Q$. Then $\{X_n\}_{n \geq 0} \preceq_{st} \{Y_n\}_{n \geq 0}$ follows from Theorem 1.

$P_{(0,0),*} \preceq_{st} Q_{(0,0),*}$ is trivially verified. For an arbitrary state $x \in \mathcal{S}^*$, $f_3(x) \in U$ implies $f_1(x), f_2(x) \in U$, for each increasing set U. For type 2 states, additionally $f_4(x) \in U$ implies $f_1(x), f_2(x) \in U$, for each increasing set U. Then from $p_1 \leq r_1$, $p_2 \leq r_2$, and Table 1 it follows that for each $x \in \mathcal{S}^*$, $\sum_{k \in U} P_{x,k} \leq \sum_{k \in U} Q_{x,k}$ for all increasing sets U. Thus, by Definition 3, $P \preceq_{st} Q$. □

Let us denote by T_X and T_Y end to end delays for DTMCs $\{X_n\}_{n \geq 0}$ and $\{Y_n\}_{n \geq 0}$.

Corollary 1. *Under the same hypothesis as in Theorem 3, $T_X \preceq_{st} T_Y$ ($\preceq_{st}$ is seen here as stochastic comparison under the usual total order on $\mathbb{N}_0$).*

Proof. Theorem 3 implies $X_n \preceq_{st} Y_n, \forall n$. In particular, for increasing set $\mathcal{S}^*$, we have $P(X_n \in \mathcal{S}^*) \leq P(Y_n \in \mathcal{S}^*), \forall n$. This gives $P(T_X \leq n) = P(X_n = (0,0)) = 1 - P(X_n \in \mathcal{S}^*) \geq 1 - P(Y_n \in \mathcal{S}^*) = P(T_Y \leq n), \forall n$. Thus, $T_X \preceq_{st} T_Y$. □

It is important to remark here that we prove an inequality on distributions of the end to end delay. This is much more useful than a comparison of expectations. For instance we can derive very easily a bound on the tail probability of the delay. Furthermore, we do not assume that the real model is monotone. So we can use now the theorem with conservative estimation of the deflection probabilities to obtain a bound on the end to end delay.

Even if the condition $p_1 \leq p_2$ is not necessary for the real model of the network, we now show briefly that it is true for saturation traffic when the link capacity is 1. We compute $d_1(i,j)$ and $d_2(i,j)$, the conditional probabilities of deflection for type 1 and 2 packets knowing that i type 1 and j type 2 packets are also competing, with independent arrivals (as in [3]).

Take care that the table of $d_1()$ and $d_2()$ given in [3] contains some typos. The corrected values are: $d_1(0,3) = 0$, $d_1(1,2) = 1/8$, $d_1(2,1) = 11/48$, $d_1(3,0) = 81/256$, $d_2(0,3) = 5/64$, $d_2(1,2) = 5/32$, $d_2(2,1) = 15/64$, and $d_2(3,0) = 9/32$.

The deflection probabilities p_1 and p_2 may be computed as a function of the probability u_1 that a link contains a type 1 packet. However, we can show that if $u_1 < 0.68$ then $p_1 < p_2$ under saturation traffic. Again, both real system and bound are not required to be monotone. Only one is necessary and the bounding matrix is proved to be monotone under the $p_1 \leq p_2$ assumption. If we are able to bound the deflection probabilities with two values which satisfy the constraint, this is sufficient to bound end to end delay.

5 Conclusion

We showed that the transition matrix of a packet in an odd torus with Deflection Routing is $\preceq_{st}$-monotone under the usual product partial order $\preceq_{\star}$. We then stated that end to end delay is stochastically increasing in deflection parameter. With the help of large deviation tools, these results open paths to compute bounds on the delay distribution. Bounds on the distribution delay are much more useful than expectations. For instance, delay distribution may permit to scale timers.

References

1. P. Baran. On distributed communication networks. *IEEE Transactions on Communication Systems*, CS-12:1–9, 1964.
2. A. Borrero, J. Fourneau, and F. Quessette. Packet selection in a deflection routing algorithm. In *ISCIS 2002, Orlando, USA*. CRC Press, 2002.
3. J. Fourneau and T. Czachórski. Transport time distribution for deflection routing on an odd torus. In *Europar 05, Portugal*, v. 3648 of *LNCS*, pp. 975–983. Springer Verlag, 2005.
4. P. Gravey, S. Gosselin, C. Guillemot, D. Chiaroni, N. L. Sauze, A. Jourdan, E. Dotaro, D. Barth, P. Berthomé, C. Laforest, S. Vial, T. Atmaca, G. Hébuterne, H. E. 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, 2001.
5. S. Mneimeh and F. Quessette. Minimum deflection routing algorithm, 2002. Alcatel Patent Application #135945.
6. A. Muller and D. Stoyan. *Comparison Methods for Stochastic Models and Risks.* Wiley, New York, NY, 2002.