A Study of Network Capacity under Deflection Routing Schemes

Josep Fàbrega and Xavier Muñoz*

Departament de Matemàtica Aplicada IV, Universitat Politècnica de Catalunya
Campus Nord C3, Jordi Girona 1–3, 08034 Barcelona, Spain
{matjfc,xml}@mat.upc.es

Abstract. Routing in bufferless networks can be performed without packet loss by deflecting packets when links are not available. The efficiency of this kind of protocol (deflection routing) is highly determined by the decision rule used to choose which packets have to be deflected when a conflict arises (i.e. when multiple packets contend fo a single outgoing link). As the load offered to the network increases the probability of collision becomes higher and it is to be expected that at a certain maximum offered load the network gets saturated. We present an analytical method to compute this maximum load that nodes can offer to the network under different deflection criteria.

1 Introduction

Deflection routing [1] is a routing scheme for bufferless networks based on the idea that if a packet cannot be sent through a certain link due to congestion, it is deflected through any other available one (instead of being buffered in a node queue) and rerouted to destination from the node at which the packet arrives. In this way, congestion causes packets admitted to the network to be misrouted temporarily, in contrast with traditional schemes where such packets might be buffered or dropped. This kind of protocol has been proposed, for instance, to route packets in all-optical networks because optical storage is not possible with nowadays technology [2,4,7] (Messages can only be shortly delayed by a fiber loop in order to wait for a quick processing of their headers, but cannot be buffered in queues without optical to electrical conversion.)

Many approximations have been proposed in the literature for implementing deflection routing [8]. The efficiency of the protocol is highly determined by the decision criteria used to deflect packets when collisions arise (i.e. when two packets should use the same outgoing link and one of them must be deflected). The strategies used to solve these conflicts can be divided into two categories. On one hand, those that give priority to the most disadvantaged packets in order to avoid deadlock or timeouts (and thus trying to guarantee that all packets arrive

* Research supported by the Ministry of Science and Technology, Spain, and the European Regional Development Fund (ERDF) under projects TIC-2001-2171 and TIC-2002-00155.

H. Kosch, L. Böszörményi, H. Hellwagner (Eds.): Euro-Par 2003, LNCS 2790, pp. 989–994, 2003.

to destination): MAXDIST (packets that are further away to destination have a higher priority), MAXTIME (older packets are given a higher priority), and MAXDEFL (the larger the number of times a packet has been deflected, the higher the priority). On the other hand, strategies that give preference to those packets that might arrive to destination as soon as possible. The decision criteria within this group are analogous to the preceding case: MINDIST, MINTIME and MINDEFL.

As the load offered to the network increases, deflection routing becomes less efficient (the probability of collision increases) and at a certain maximum offered load R the network becomes saturated. Figure 1 shows the results of simulations under different decision criteria for a certain network topology. As the reader can check, the maximum allowed traffic highly depends on the decision rule used to deflect packets. This paper shows an analytical method to compute this maximum load R. Some specific theoretical results for the hypercube and shufflenet networks can be found in [11,10].

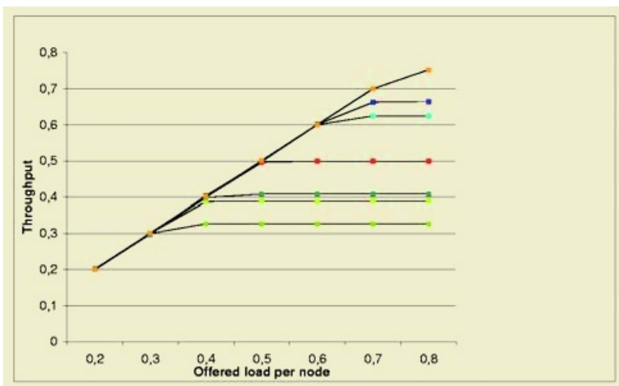

Fig. 1. Delivered vs. offered traffic

2 The Model

Network topology is another important aspect that determines the efficiency of deflection routing. We suppose here that the network is modeled by a directed graph (digraph) in which all nodes offer the same constant traffic load to the network by means of an input queue. We also assume that the communication model have the following properties: packets have fixed length and are transmitted synchronously, processing time at the intermediate nodes is zero, only one packet can travel through a given link at a time, and a packet in an input queue enters in the network as soon as a link is available and competes with other packets in the network for the same links (transmit no hold acceptance (txnh). (Other queue management policies can be found in [5].) With respect to the topology and routing table the following restrictions are also supposed: all nodes have the same number of outgoing and incoming links (i.e. the digraph is δ-regular), packets will always try to follow the shortest path to destination, and there is only one shortest path between any pair of nodes, which means that a packet remains at the same distance or further away to its destination after being deflected. Besides the previous assumptions we consider two more suppositions in order to simplify our problem. Firstly, a packet entering to a node exits through any outgoing arc with probability $1/\delta$ (uniform traffic distribution). This approximation is reasonable if the routing table is such that the edge-forwarding index [12] is close to the minimum. Secondly, since the purpose of our study is to deal under heavy traffic conditions, it will be assumed that each node has a packet ready to be transmitted as soon as a link is free. This supposition holds only for deterministic distribution of packet arrivals but it is not true for other traffic distributions such as Poisson or sporadic. Nevertheless the supposition is acceptable if the arrival ratio is high enough and if there are large input buffers which are overflown.

With these assumptions, in steady-state the number of incoming packets to a node equals the number of incoming links to that node, and equals also the number of outgoing packets and outgoing links. In other words, the probability of a link being free during a time unit is zero. Hence the number of packets in the network equals de number of arcs in the digraph, and then, by Little's Law we have $nR\bar{t} = m$ where R stands for the maximum admissible throughput per node (i.e. the maximum load rate nodes can offer to the network) and $\bar{t}$ is the average time for a packet to arrive to destination. In the case of δ–regular directed graphs, we have $R = \delta/\bar{t}$. The following two sections will be devoted to give methods to compute R under different decision criteria.

3 Random Policy

To compute the average time $\bar{t}$ that (in steady-state) a packet is in the network hopping through its nodes, let us define an absorbing Markov chain with states corresponding to the possible distances that the packet could be to its destination: $0, 1, \ldots, D$, where D is the diameter of the digraph and state zero stands for a packet that has arrived at destination (see Figure 2). The transition

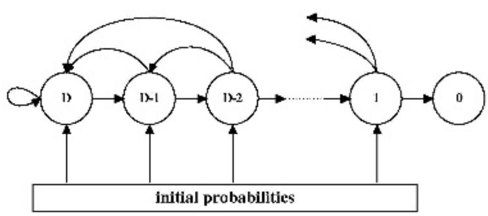

Fig. 2. Markov chain used to compute the average time to arrive to destination

probability matrix $M = (m_{ij})$ of this chain is given by $m_{ij} = P_a(i) = 1 - P_d(i)$ if $j = i-1$ and $i > 0$, $m_{ij} = P_d(i)P_t(i,j)$ if $j \geq i$ and $i > 0$, and 0 otherwise, where $P_d(i)$ is the probability that a packet being at distance i is deflected and $P_t(i,j)$ is the conditional probability that a deflected packet which is at distance i to its destination z goes to a node which is at distance j to z.

The deflection probability $P_d(i)$ depends on the number of packets that should use the same link at a time. Given a packet that wants to use certain link, the probability that N other packets want to use the same outgoing link (N collisions) is given by $P_c(N) = \binom{\delta-1}{N}(1-1/\delta)^{\delta-1-N}(1/\delta)^N$. Therefore, the probability of being deflected conditioned to being at distance i to destination is $P_d(i) = \sum_{N=1}^{\delta-1}(1 - P_a(i \mid N))\ P_c(N)$, where $P_a(i \mid N)$ stands for the probability $P_a(i)$ conditioned to N collisions. In the case of *random policy*, this probability does not depend on i, and according to assumption of uniform traffic, is given by $P_a(i \mid N) = 1/(N+1)$. Consequently, in the case of random policy, the probability of deflection is, for any distance i, given by $P_d = \sum_{N=1}^{\delta-1}(1 - 1/(N+1))\binom{\delta-1}{N}(1-1/\delta)^{\delta-1-N}(1/\delta)^N = (1-1/\delta)^\delta$.

On the other hand, the probabilities $P_t(i,j)$ depend only on the network topology as well as the initial probability $P_{in}(i)$ of each state (i.e. the probability that a new packet entering the network is assigned a destination node at distance i from the source). A detailed analysis of these probabilities can be found in [9] for the case of Kautz networks [3].

Once defined the transition and initial probabilities, the Markov chain makes it possible to compute the average time that a packet is in the network by computing the mean time $\bar{t}$ to absorption to the zero state. This time will give us the maximum admissible throughput $R = \delta/\bar{t}$.

4 Distance Priority Criteria

In case of distance criteria, the probability that a packet advances towards destination depends on its distance to that destination, but also on the probability that competitors are closer to or further to its own destinations. In other words, to compute the probability of advancing $P_a(i) = 1 - P_d(i)$ we need to know the probability of being at each state in the Markov chain. Hence, a more complex analysis must be performed.

Since packets fighting for a certain outgoing link have not arrived yet to destination (i.e. they are not in the zero state) we can consider the probabilities of being at each state conditioned to not being in state 0. Let us call these new conditioned state probabilities $P'(i)$. In order to compute $P'(i)$ and the new transition probabilities it is convenient to state the problem in a different way: there is a fixed number of packets (equal to the number of links) hopping through the nodes in such a way that they will never exit the network, but once a packet reaches its destination, a new random destination is immediately assigned to it. The new ergodic Markov chain associated to this problem is shown in Figure 3. Because of the assumption that each node has a packet ready to be transmitted as soon as a link is free, the traffic in the network will be exactly the same as in

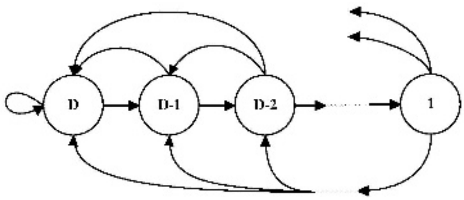

Fig. 3. Conditioned Markov chain

the original problem and, moreover, the transition probabilities will also be the same for $i \geq 2$. More precisely, the entries m'_{ij} of the new transition probability matrix $M' = (m'_{ij})$ are $1 - P_d(i)$ if $j = i - 1$ and $i \neq 1$, $P_d(i)P_t(i,j)$ if $j \geq i$ and $i \neq 1$, $P_d(i)P_t(i,j) + (1 - P_d(i))P_{in}(j)$ if $i = 1$, and 0 otherwise, where $P_{in}(j)$ stands for the probability that the new assigned destination is at distance j from source, as in the preceding section.

Each deflection probability $P_d(i)$ is now a function of the state probabilities $P'(1), \ldots, P'(D)$. For instance, if MINDIST criteria is used, $P_d(i)$ can be computed as in the previous section, but now with the advancing probability $P_a(i|N) = \sum_{k=0}^{N} \binom{N}{k} (P'(d > i)^{N-k} P'(i)^k)/(k+1)$ if $i < D$ and $P_a(i|N) = P'(D)^N/(N+1)$ if $i = D$, where $P'(d > i)$ stands for the probability that a competitor packet is at distance greater than i to its destination.

Table 1. Results from theory and simulation

		MINDIST		MAXDIST		Random policy	
d	D	Theor.	Simul.	Theor.	Simul.	Theor.	Simul.
2	3	0,52	0,54	0,37	0,39	0,46	0,46
2	4	0,37	0,38	0,23	0,22	0,30	0,30
3	3	0,62	0,69	0,39	0,41	0,54	0,55
3	4	0,43	0,47	0,21	0,20	0,33	0,33
4	3	0,75	$> 0,80$	0,45	0,45	0,64	0,65
4	4	0,53	0,58	0,22	0,20	0,39	0,38

Even if the transition probabilities are unknown, M' is the probability matrix of an ergodic Markov chain (finite, irreducible and aperiodic). Let $\mathbf{V}' = (V'(1), \ldots, V'(D))$ be the stationary distribution of M' (i.e. a probability left eigenvector associated to the eigenvalue 1), each $V'(i)$ being a function of $\mathbf{P}' = (P'(1), \ldots, P'(D))$. The probabilities $P_d(i)$ can be computed then by solving the equation $\mathbf{V}' = \mathbf{P}'$. Finally, we consider again the absorbing Markov chain M and compute the expected time to absorption in state 0, as in Section 3.

5 Validation through Simulation

Table 1 gives, for the case of the d-regular Kautz network $K(d, D)$ with diameter D, a comparison of theoretical results obtained by applying our analytical model with results from simulation.

In the computation of the theoretical results, the values of $P_{in}(i)$ and $P_t(i, j)$ are those obtained in [9]. The comparison is a validation of the hypothesis and assumptions used in paper. Further details and description of the simulation can be found in [14].

References

1. Baran, P.: On distributed communications networks. IEEE Trans. Comm. Sys. **12** (1964) 1–9
2. Baransel, C., Dobosiewicz, W., Gburzynski, P.: Routing in multihop packet switching: Gb/s challenge. IEEE Network **9** (1995) 38–61
3. Bermond, J.C.; Peyrat, C.: De Bruijn and Kautz networks: A competitor for the hypercube? In: Andre, F., Verjus, J.P. (eds): Hypercube and Distributed Computers, North-Holland, Amsterdam (1989) 279–294.
4. Borgonovo, F., Fratta, L.: Deflection networks: Architectures for metropolitan and wide area networks. Comput. Networks ISDN Syst. **2** (1992) 171–183
5. Bononi, A., Pruncal, P.R.: Analytical evaluation of improved access techniques in deflection routing networks. IEEE/ACM Trans. on Networking **4(5)** (1996) 726–730
6. Chich, T.: Optimisation du routage par déflexion dans les réseaux de télecommunications métropolitains. Ph.D. Thesis. LRI Paris (1997)
7. Chich, T.; Cohen, J.; Fraigniaud, P.: Unslotted deflection routing: a practical and efficient protocol for multihop optical networks. IEEE/ACM Trans. on Networking **9** (2001) 47–59
8. Clérot., F.: Réseaux fonctionnant par déflexion: deux ans déjà... Internal Research Rapport. CNET (1996)
9. Fàbrega, J.; Muñoz,X.: How to find the maximum admisible throughput in deflection routing networks. Internal Research Rapport. MA4. UPC (2002)
10. Gary, S.-H., Kobayashi, H.: Performance analysis of Shufflenet with deflection routing. Proceedings of IEEE Globecom 93 **2** (1993) 854–859
11. Greenberg, A.G.; Hajek, B.: Deflection routing in hypercube networks. IEEE Trans. Comm. **40** (1992) 1070–1081
12. Heydemann, M.-C., Meyer,J. C., Sotteau, D.: On forwarding indices of networks. Disc. Appl. Math. **23** (1989) 103–123
13. Homobono, N.: Resisstance aux pannes des grands réseaux d'interconnexion. Ph.D. Thesis. LRI Paris (1987)
14. Ortega. D.: Simulación de encaminamientos por deflexión sobre algunas familias de digrafos. Master Thesis. UPC. Barcelona (2000)