Fault Tolerance in Optical Networks; A study of electronic in- and egress interconnections in Torus topologies

Bjarne E Helvik[1], Ragnar Ø Andreassen[2]

[1]Norwegian University of Science and Technology [Bjarne.E.Helvik@item.ntnu.no]
[2]Telenor R&D [ragnar-oivind.andreassen@telenor.com]

Abstract

The ingress and egress nodes in an all-optical network have an electronic in- and egress part and an optical switch. This work analyses various architectures for interconnection of the "electronic" and "optical" parts of the network in order to tolerate single optical switch failures and single optical link failures. The architectures studied are: non-redundant (reference), fault tolerant optical nodes and multihoming of the electronic in- and egress to home node and all neighbouring nodes. The network topology studied is a two-dimensional torus network with symmetric traffic. Network configurations having comparable properties with regard to dependability and traffic transport capacity are compared with respect to link capacity and node switching requirements, i.e. cost.

The analysis shows that the fault tolerant networks have nearly the same requirement for link redundancy. The multihomed architecture has less optical switching requirements than even the non-redundant case, and is definitely the most attractive solution with respect to the proposed evaluation criteria. The switching requirement of the fault-tolerant node architecture, where duplication is assumed to tolerate optical switch faults, is far the demanding with respect switching cost/requirements.

1 Introduction

In order to ensure robustness with respect to failures, emerging optical core networks will need to be designed with redundancy. A basic fault tolerance design question is whether to build redundancy into optical nodes, making them inherently fault tolerant, or to give these nodes a simple design and provide fault tolerance by other means. Considering that optical networking technologies, and in particular optical burst and packet switching technologies, are still immature, the cost efficient alternative may be to choose the latter approach. The objective of this paper is to pursue an architecture where optical nodes are kept simple, and where single link or node failures are tolerated either by switch duplication or access multihoming [1]. The particular multihoming solution pursued is where the "electronic" part of the ingress-egress nodes in the optical network is multihomed to neighbouring optical switches in addition to being connected to the local optical switch. This electronic multihomed part of the node is in the following referred to as EIE (Electronic In- and Egress). The quantitative results obtained are independent of the switching technology and may be applied for circuit and packet switched networks.

The basic quantity considered here is an aggregate unidirectional end-to-end stream. Hence, if k node pairs communicate via a link or node in the network the "load" of the link is k. Dependent on the offered traffic between pairs of nodes and the switching technology employed, the number of paths between nodes may be converted into a physical capacity requirement, cf. [2]. The requirement for physical resources, e.g. number of fibres and wavelengths, may then be derived from the capacity requirement. Required link capacity (in terms of end-to-end stream transport capacity) and optical switch path terminations are denoted $k(M)$ and $s(M)$ respectively, where M is the number of nodes in the network. Both the switch duplication and multihoming solutions investigated will tolerate node or single link unavailability situations.

In this analysis, the differences in capacity requirements due to different resource sharing models (dedicated versus dynamically shared) are not considered. This second order effect may be introduced at a later stage following the model presented in [2].

2 Network

2.1 The network

As in [2] and [3] an $M = m^2$ torus network with a homogeneous offered traffic load is regarded to obtain parametric results. To limit the mathematical complexity, only odd values of μ are considered. The 2D torus, see the figure below, is considered as the symmetric topology that best represents a core network.

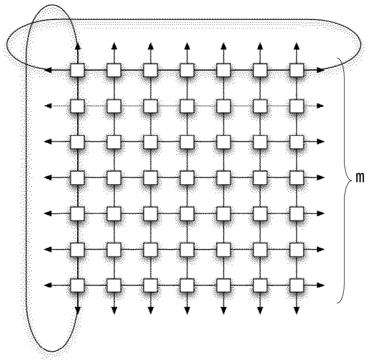

Figure 1 Symmetrical $M = \mathtt{m}^2$ torus network topology. Each line in the figure represents two oppositely directed unidirectional links.

Each node in the network, cf. [1] section 2.6, is separated in an electronic ingress/ egress part and an optical switching part, as illustrated in the figure below.

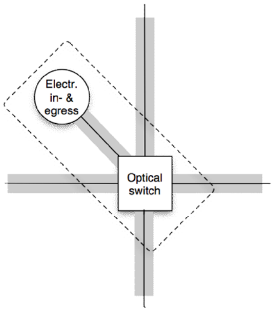

Figure 2 Divisioning of a network node into an electronic in- and egress part and an optical switch.

The routing of traffic in the network is done according to the shortest path with minimum number of bends, i,e., none when source and destination is on the same row or column and one otherwise. There is one path from each source to each destination. The bend is always done in the same direction, e.g., clockwise, and since the traffic is symmetric, there will be the same number of paths on each link.

To perform a comparison of various network architectures, cost models of the network elements, i.e. optical nodes and links between them are needed. It is outside the scope of this work to establish such models. All capacity requirements will here thus be computed in the more abstract quantities $k(M)$ and $s(M)$.

3 Capacity requirements

3.1 Non fault-tolerant architecture

In the non fault-tolerant architecture any link or node failure will result in loss of the traffic carried by the network element. This case will serve as a reference in the following. The detailed node interconnection structure is as depicted in Figure 2 above.

To calculate the number of node-to-node paths on each link, consider the network depicted in Figure 3, where (unidirectional) paths are routed according to a shortest path, minimum number of clockwise bends strategy. We consider the paths from one node to all other nodes in the network. Due to symmetry, we can focus on the central node in the network without loss in generality.

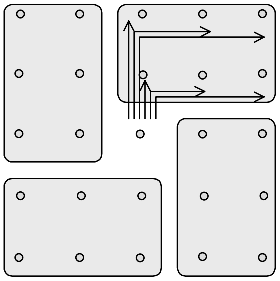

Figure 3 Calculation of paths per link in the non-redundant case

The number of hops for links from the central node to other nodes in a shaded sector can be found by summing up the lengths of paths for each row of the sector. Multiplying by the number of sectors per node (four) and number of nodes (M), and dividing by the total number of links ($4M$) will give the number of paths per link k:

$$k = 4M\left(\sum_{i=1}^{(\mu-1)/2}\sum_{j=0}^{(\mu-1)/2} i + j\right)/4M = \sqrt{M}(M-1)/8$$

The number k represents the minimum number of divisions of a physical link resource in the network we are considering. It was derived here under a particular routing scheme, but is invariant to all symmetrical shortest path routing schemes [2]. The number k also represents the requirement to traffic transport capability of link resources in terms of unit end-to-end flows. An operator might choose to split traffic between end points into different, possibly disjoint, paths. As long as the network symmetry is maintained, the transport capability requirement of links is still *k*. Depending on the technology, such divisioning may however have impact on the actual physical link capacities needed in order to convey the overall traffic [2].

There is one (aggregate) unit end-to-end stream from each electronic in- and egress (EIE) node to each of the others, in total *M*-1. Hence, the termination capacity in terms of unit end-to-end streams of the optical switch that determines its size and cost becomes:

$$s = M - 1 + 4k(M) = (M-1)\left(1 + \sqrt{M}/2\right)$$

3.2 Fault tolerant nodes

We consider the case where the nodes (optical switches) are made fault tolerant by duplication. The link capacity within the network is increased so a single link failure will not cause any traffic loss. The traffic is routed as follows: traffic from a node to another node not on the same row or column is split evenly between two shortest, minimum number of bends, disjoint paths; one in the clockwise and one in the anti-clockwise direction. Upon link outage, the traffic in the affected path is rerouted to the opposite path. Between nodes in the same row or column, traffic is sent directly in the shortest direction under normal operation. Upon failure on this stretch, traffic is routed by the shortest alternative of either the opposite direction, or split between two new disjoint shortest, minimum number of bends paths.

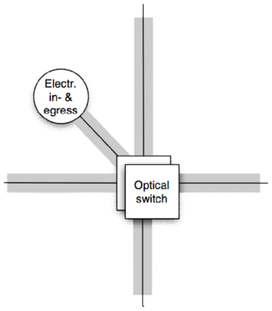

Figure 4 Node architecture with fault tolerant optical switches

In order for the paper to be reasonably self-contained, we give an outline of the calculation of the capacity required to tolerate single link failures. We will refer to the network as indicated in the figure below.

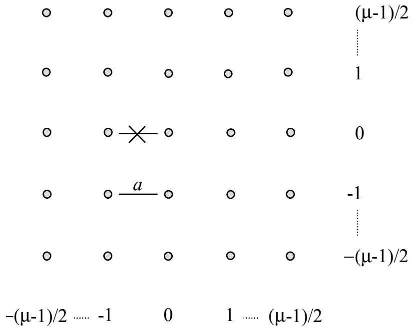

Figure 5 Network with failed link at location [(0, 0), (-1, 0)]

We number rows and columns as indicated. We shall use a coordinate system convention here and in the following to describe positions in the network. Following [3] we define the sets of nodes {*A*} and {*B*}, containing the nodes in row zero and the nodes not in row zero respectively. For the routing scheme chosen, traffic between a pair of nodes in set *B* will not be affected by the outage, so we need only consider the traffic either inside *A* or originating in *A* and terminating in *B* or vice versa. For the traffic on links perpendicular to the failed link, the following apply: There will be no additional traffic from traffic between pairs of nodes in different sets *A* and *B*. As shown in [3], node pairs can in this case be found such that increased traffic from one node on a perpendicular link due to rerouting is compensated with the decreased traffic due to rerouting from the other node in the pair on that link. There *will* be traffic increase on perpendicular links from traffic rerouted between some nodes within *A*. (Those for which rerouting in the same direction is shorter than rerouting in the opposite direction.) This traffic will also pass parallel links (in particular link *a*), which will have to accommodate additional rerouted traffic. The traffic increase on parallel links will therefore be greater than that on perpendicular links.

We now consider the additional traffic on the link *a* drawn in position [(-1,-1),(0,-1)]. This will consist of two components; traffic rerouted between nodes in rows 0 and –1, and traffic rerouted between nodes within row 0.

For the first part, consider the traffic from column 0 headed leftwards. The traffic from the two relevant nodes ((0,0) and (0,-1)) in that direction is destined for nodes leftwards to column -(μ-1)/2. Both nodes will increase the load on link *a* due to redirection with a half unit to each of the (μ-1)/2 destination nodes. From the two corresponding nodes in column 1 the number of nodes to be reached through link *a* will be (μ-1)/2 –1 and so forth for the rest of the (μ-1)/2 columns in the right half of the network. The traffic increase for the first part will therefore be $\sum_{j=1}^{(\mu-1)/2}(j/2+j/2)$. Considering now the rerouting of traffic within the row 0, it can be shown [3] that the nodes up to a distance (μ-3)/2 from the failed link will reroute traffic on the two parallel links in the same direction while the remaining nodes will reroute traffic on the same row in the opposite direction. Following a similar logic as above, the contribution of that component to link *a* will be $\sum_{j=1}^{(\mu-3)/2} j/2$. It can be shown ([3] again), that the link *a* will receive the maximum additional load in the network. Adding the two components above and the original traffic thus yields the required capacity to handle single link failures without loss:

$$k_{ft} = k + \sum_{j=1}^{(\mu-1)/2} j + \sum_{j=1}^{(\mu-3)/2} j/2 = \frac{1}{16}\left(\sqrt{M}-1\right)\left(2M+5\sqrt{M}-1\right).$$

The corresponding termination capacity of optical switches can be calculated as for the non-fault tolerant case, only catering for the node duplication by doubling the requirement for termination capacity.

$$s_{ft} = 2(M-1+4k_{ft}) = \frac{1}{2}\left[\sqrt{M}\left(2M+7\sqrt{M}-6\right)-3\right]$$

3.3 Multihoming

In this configuration, see Figure 6; optical node failures are tolerated by multihoming the electronic in- and egress part of a node (EIE) to the optical switches of its four neighbouring nodes as well as to the home node. The traffic from the EIE in fault free operation is routed directly to the neighbouring nodes. Similarly for terminating traffic, with the exception of traffic from neighbouring nodes which is terminated in the destination EIEs local optical mode.

As for the fault tolerant node case, unidirectional traffic from one node to another is bifurcated between two disjoint shortest– minimum number of bends paths. An outline of the calculation of the link capacity needed in order to handle single node failures according to the above redundancy scheme will be given, using Figure 7 below as a reference.

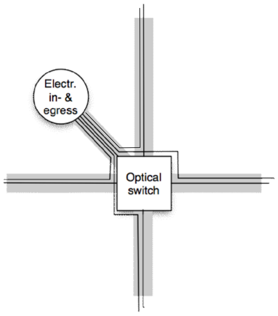

Figure 6 Node interconnection architecture in the multihoming case

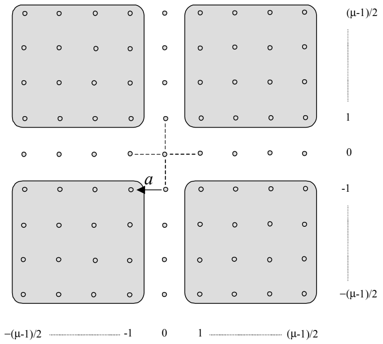

Figure 7 Network with failed node

We consider the central node of the network that will be representative due to symmetry. It can be shown, following the reasoning in [3], that the links surrounding the failed node will receive the greatest additional load. We thus concentrate on additional traffic occurring on link *a* in the direction as indicated on the figure. All other links in the inner square are receiving the same additional load due to the rotational and mirroring symmetries of the network.

It is immediately evident that traffic between, or internal to, the shaded regions in the figure will not cause additional load on *a*. We thus need only to consider traffic between a shaded region and a white region, or internal to the white region. This traffic can be divided into six groups:

- Traffic from nodes in row –1 destined for nodes in row 0. Half of this should have gone over the failed 0-row links, giving an extra contribution to *a* of $\sum_{j=1}^{(\mu-1)/2} j/2$ units.
- Traffic from the (μ-5)/2 nodes in columns 1 and rightwards in row 0 which should have passed origo to the opposite nodes. Half of this traffic will pass through *a*, giving the contribution $\sum_{j=1}^{(\mu-5)/2} j/2$ units.
- Traffic from node at (0, -1) destined for the upper left shaded region, giving the contribution of $(1/2)[(\mu-1)/2][(\mu-3)/2]$ units.
- Traffic from node at (0, -1) destined for the (μ-5)/2 nodes in column 0 from rows 1 and upwards, giving the contribution [(μ-5)/2]/2 units to *a*.
- Traffic from row 0 to row –1, giving a contribution of $\sum_{j=1}^{(\mu-3)/2} j/2$.
- The contribution from the above nodes is cancelled out by reduced load from nodes in column zero from traffic which should have passed *a* on the way to nodes in row –1, but which will be rerouted outside *a* upon failure.

The multihoming case must also withstand single link failures. As the sum of the above contributions totals less than the single link failure redundancy for small networks, we get the following expression for the required link capacity in the multihomed case:

$$k_{mh} = Max\left[k + \sum_{j=1}^{(\mu-1)/2} j/2 + \sum_{j=1}^{(\mu-5)/2} j/2 + \frac{(\mu-1)(\mu-3)}{8} + \frac{\mu-5}{4},\ k_{ft}\right],$$

giving:

$$k_{mh} = Max\left[\frac{1}{8}\sqrt{M}\left(M + 2\sqrt{M} - 7\right),\ k_{ft}\right].$$

Note that we in the above calculations have not discriminated between traffic flowing on the direct multihoming links and the links between switches, but rather considered the total. One possible dimensioning strategy would be to keep capacity at the minimum level at the multihoming links (i.e. (*M*-1)/4), and take out the increase on the links between switches.

There will be a reduction in switch link termination requirements in the multihoming case due to the direct traffic between EIEs and neighbouring nodes corresponding to the capacity carried on the direct links. The direct link between an EIE and its home switch will under ordinary operation carry four streams (from neighbouring nodes). In the single direct link failure case, the total capacity that need to be carried on the local EIE-switch link will thus be (*M*-1)/4 + 3.

Considering the above, the node-switching requirement in the multihoming case will be:

$$s_{mh} = 4[k_{mh} - (M-1)/4] + (M-1)/4 + 3 = \left(\sqrt{M}(2M + \sqrt{M} - 14) + 15\right)/4\,.$$

4 Comparison of architectures

4.1 Link requirements

Results for the link requirements of the various network architectures are presented in Figure 8 and Figure 9 below.

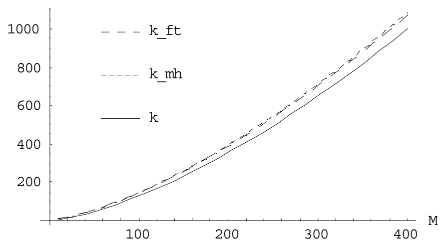

Figure 8 Link capacities required to tolerate single link or node unavailability in the cases of fault tolerant nodes (k_{ft}) and multihoming (k_{mh}) as a function of network size (*M*) in terms of unit end-to-end streams (including values for the reference non-fault tolerant case, (*k*).

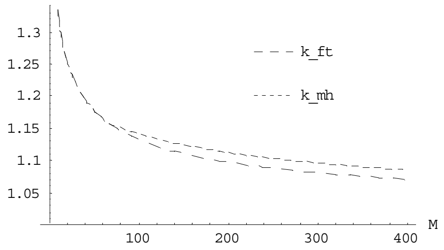

Figure 9 Corresponding capacities relative to the non-fault tolerant case.

As we can observe, the additional link capacity in terms of unit end-to-end streams required in order to tolerate node or single link unavailability is moderate for both the multihoming and the fault tolerant node case. The two alternatives differ most notably for the two smallest network alternatives, corresponding to 9 and 25 nodes respectively. For large networks, the term of $M^{3/2}$ present in all link requirement expressions will dominate, ensuring convergence to unity in relative link redundancy for both multihoming and the fault tolerant node networks.

4.2 Switching requirements

Results for the switching requirements of the various network architectures are presented in Figure 10 and Figure 11. As an additional reference, we have included plots also for the case where single links failures only are tolerated. It is seen that the multihomed architecture in fact has a lower switching requirement than the non-redundant architecture. This is due to the fewer optical switches handling an in-to-egress path.

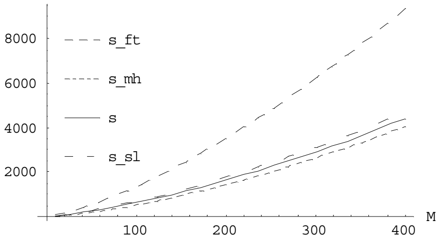

Figure 10 Node switching capacities required to tolerate single link or node unavailability in the cases of fault tolerant nodes (s_{ft}) and multihoming (s_{mh}) as a function of network size (M) in terms of unit end-to-end streams (including values for the reference non-fault tolerant (s). and single link fault tolerant (s_{sl}).cases.

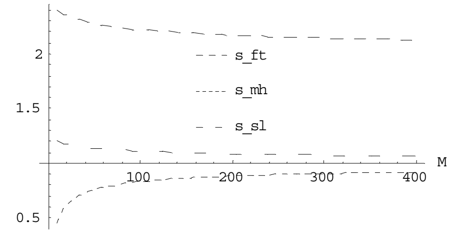

Figure 11 Corresponding capacities relative to the non-fault tolerant case.

The relative switching requirement shows that for small networks, the multihomed networks has a very small requirement, while the requirement of the fault tolerant node architecture is large even without considering the dual-factor.

5 Concluding remarks

We have studied two alternatives for ensuring fault tolerance in optical core networks. Both architectures studied will tolerate single link and single node failures. Network link- and switching requirements were computed in terms of end-to-end unit streams transport capacity. A basic assumption in the work is that such core networks are well represented by symmetrical torus-networks. In support of that assumption one can argue that the torus has a similar connectivity as one will experience in more realistic alternatives.

Both architectures studied have fairly similar link transport requirements. The fault tolerant node architecture will however have a switching requirement (node cost) far larger than the both the non-fault tolerant and the multihoming architectures. As the multihomed architecture has a lower optical switching requirement than even a non-fault tolerant network, it is the dominant least expensive (best) of the architectures investigated. This may of course be nuanced in a practical implementation due to a reduced flexibility and a division of transmission capacity into dedicated resource pools.

6 Acknowledgements

This work has been funded in part by the European Commission, IST project no. 2000-28557, STOLAS. The authors gratefully acknowledge the collaboration with project colleagues within Telenor.

7 References

[1] Node and Network Architectures and Evaluation of Network Performance, IST-2000-28557 STOLAS Deliverable D121, 31. March 2003

[2] R Ø Andreassen, B E Helvik, "A note on traffic performance gain for optical burst/packet switching", Proc. Optical Network Design & Modelling Conference, ONDM 2004.

[3] A.S. Acampora, R.A. Gholmieht, S. Krishnamurthy, On Tolerating Single Link, Double Link, and Nodal Failures in Symmetric Grid Networks, Wireless communication and Networking conference (WCNC) 2000, pp. 572-581