Telecommun Syst
DOI 10.1007/s11235-009-9277-3

Survivability and reliability of a composite-star transport network with disconnected core switches

Stefano Secci · Brunilde Sansò

Abstract This paper deals with the design and dimensioning of a novel survivable optical network structure, called Petaweb, that can reach a total capacity of several Pb/s (10^{15} bit/s). The Petaweb has a composite-star architecture that allows two-hop connections between edge nodes through disconnected core nodes. Prior studies of the same authors have tackled the optimization of a Petaweb network architecture with regular and quasi-regular topologies. In this paper, reliability and survivability issues are addressed by introducing a dedicated path protection strategy into the design model. We present by extensive numerical results the reliability and survivability properties of the Petaweb core architecture with respect to single fiber link, core node, or switching plane failure and to switching site disconnection.

Keywords Petaweb · Composite star network · Dedicated path protection · Survivability · Reliability · TDM/WDM

1 Introduction

The Petaweb [1–3] is a novel architecture with a composite-star physical topology that significantly simplifies traffic engineering functions such as routing and addressing. It offers easily-manageable and independently-configured core nodes since all its components are modular and can be extended without corrupting the existing equipment [4]. Furthermore, given the regularity of the structure, an upgrade will not jeopardize the management of an optimized network as idle capacity can be easily allocated without compromising network management.

The first Petaweb emulation was conducted by Blouin et al. [5]. Other studies have compared the Petaweb to typical optical mesh networks [4] concluding that even if the Petaweb demands a more important quantity of optical fibers (roughly 17% more than in a multi-hop network), the simplifications in traffic engineering procedures, and the simple linear extensibility, are very significant. The authors observe that the decreasing cost of optical fibers encourages the installation of this type of network. Moreover, the adoption of a quasi-regular topology [6] demonstrates that the fiber costs can be kept between 60% and 73% of the total network cost. The Petaweb has also been considered as a construction block for a structure capable of operating at the Yotta-bit per seconds (10^{24} bps) called the YottaWeb; architectural rules and optimization procedures for regular and irregular YottaWeb topologies have been presented in [7].

Despite the numerous advantages of the structure, there is, however, one critical drawback: a trunk failure may cause the total isolation of an edge node if all the enabled nodes are placed in the same site. This is particularly important given that the Petaweb is capable of operating at rates of the order of petabits per seconds, and a single edge node may represent an entire city.

The objective of this work is therefore to deal with these survivability issues of Petaweb networks. We tackle the optimization of a Petaweb network architecture with a Dedicated Path Protection (DPP) strategy; we want to avoid

Note: Preliminary results for the contents of this paper have been presented at the NAEC 2006 international conference.

S. Secci (✉)
Institut Telecom, Telecom ParisTech, LTCI CNRS, 37/39, rue Dareau, 75014 Paris, France
e-mail: secci@telecom-paristech.fr

B. Sansò
GERAD and Department of Electrical Engineering, École Polytechnique de Montréal, C.P. 6079, Succ. Centre-ville, Montréal, Québec, Canada H3C 3A7
e-mail: brunilde.sanso@polymtl.ca

Published online: 22 January 2010

the design of Petaweb physical topologies where restoration functions cannot allocate dynamically the protection paths because of the possible isolation of edge nodes. The paper is structured as follows. The network model is described in Sect. 2. Section 3 deals with the protection strategy. The design problem and the proposed resolution method are discussed in Sect. 4. Network dimensioning results are presented and discussed in Sect. 5 whereas performance and robustness evaluation results are discussed in Sect. 6. Section 7 contains conclusions and suggestions for further work.

2 Network model

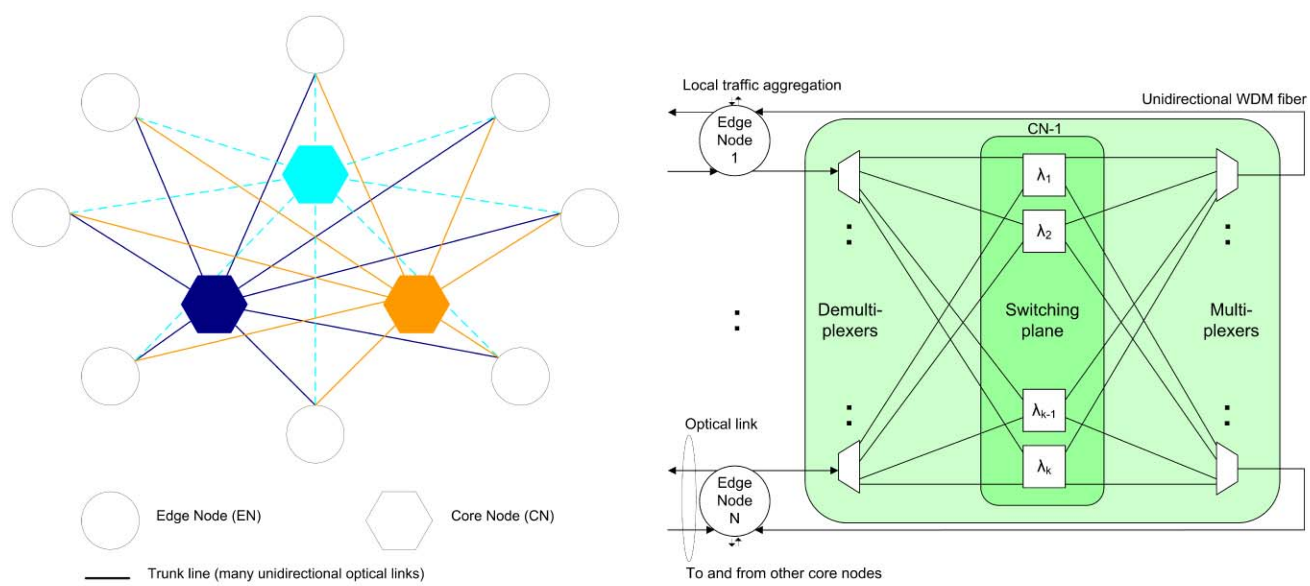
The Petaweb composite-star architecture is depicted in Fig. 1. Every edge node (EN) is connected to every core node (CN) but neither the core nodes nor the edge nodes are directly connected to each other. This kind of configuration allows a two-hop optical path between edge nodes through a single core node, thus reducing routing and addressing complexity and facilitating upgrades and expansions due to its regularity.

To our knowledge, the Petaweb is the only architecture that presents disconnected *backbone* nodes.

In this paper, we assume that a Connection Request (CR) consists of traffic coming from a SONET/SDH interface, that the connection between an EN and a CN is done through an optical link composed of one or more fibers and that the EN has one optical link ingress and one egress for every core node it is associated with.

A *core node* (CN) is made up of arrays of space switches or *switching planes*. Let s_r be the number of switching planes in a core node and CN_r a core node with s_r switching planes; such a core node will be referred to as of type r. Three types of core nodes are defined: with one, two and four switching arrays, respectively. Figure 1b shows a type 1 array. All the ingress WDM fibers are demultiplexed into their different lambda-channels, each of which is connected to the corresponding space switch of the corresponding array. With this switching node no wavelength conversion is needed, and no wavelength continuity constraint needs to be applied. Each space switch handles channels of the same wavelength; those referred to the same EN are then multiplexed into the optical link going back to that EN. Note that due to the parallel planes, a failure in a switching plane only affects that particular plane, increasing CN reliability. Also note that a single fiber sends or receives traffic to or from a single switching plane. The network design problem is to find the best CN location and type that minimizes a cost function. Note that when an EN and a CN are co-located, the traffic from the EN is directly added and dropped. The physical connection between an EN and a switching site is called *optical trunk line* and since there could be several CNs in the same site, several optical links can compose a single trunk line.

The objective is to minimize the total network cost that is composed of the core node cost, the fiber cost and the propagation delay cost. The *CN cost* is composed of a fixed cost f_r and the cost of the ports. f_r depends on the type r of the CN such that $f_r > f_{r-1} > \cdots > f_1$. Note also that the number of switching planes is such that $s_r = 2s_{r-1}$ ($s_1 = 1$). Concerning the cost of the ports, we consider that an active port has a cost P scaled for higher types. Let N_{en} be the



(a) Composite-star physical topology

(b) Parallel-planes optical core node

Fig. 1 The Petaweb Architecture

number of ENs and γ the scale factor for P, the global cost of a CN-r is $K_r = f_r + 2N_{en} W s_r P \gamma^{(s_r - 1)}$, such that $K_r < 2K_{r-1}$ because the second cost term of K_r is scaled by γ for CN-2 and by γ^3 for CN-3.

Let $F_{i,r}$ be the total fiber installation cost due to the placement of a CN-r at site i. It can be generally expressed as $F_{i,r} = \sum_j c(i, j, r)$, where $c(i, j, r)$ represents the fiber cost to connect site j to a CN-r in site i. It may include the cost of leased lines or of new lines. For instance, if a leased solution is the least cost solution between sites i' and j', then the $c(i', j', r)$ cost contribution would represent the lease rate plus the cross connect charges; or, if the establishment of new fiber links is required, it would reflect also the trenching cost, the cost of fibers with amplifiers and/or regenerators, cross connect charges, etc.

Let β be the *propagation delay cost*, which is proportional to the distance traveled and to the lightpath bitrate. Such a cost, that was first added in [3, 6, 8] prevents a too large lightpath length, thus reducing link failure probability and propagation delay.

2.1 Classes of service

We introduce in this section the notion of classes of services to be able to accommodate the traffic in the physical channels. As adopted in [4, 6] that introduce Petaweb TDM, let us call a time-slotted lightpath, a *ts-lightpath* (TLP).

We introduce a three-level hierarchy for ts-lightpaths inspired by the OTU hierarchy such as the one proposed in [9], but without limiting our choice of bit-rate values to the OTU rates. Let Z_h be the transport capacity of a TLP of class h (TLP-h). In this paper we define the following capacity granularities that further simply the switching:

- $Z_1 = \frac{1}{2^n} C_{ch}$, $n \in \mathbf{N}$, represents the transport capacity of a time-slot; it is a fraction, multiple of 2, of the transport capacity of a wavelength (C_{ch});
- $Z_2 = C_{ch}$, that is the transport capacity of a wavelength
- $Z_3 = W C_{ch}$, that is the transport capacity of a fiber

A core node can commute a TLP-1 in a time-slot basis, a TLP-2 switching the whole wavelength without time-slot alignment, and TLP-3 by simple interconnecting the ingress fiber with the egress fiber, hence without any demultiplexing/multiplexing. The original data flow fragmented into TLPs at the source EN is then recomposed at the destination EN. We assume $C_{ch} = 10$ Gb/s, $n = 4$ and $W = 16$, hence $Z_2 = 10$ Gb/s, $Z_3 = 160$ Gb/s and $Z_1 = 0.625$ Gb/s (this last value has also been chosen to obtain the best efficiency for the given traffic volumes). These bit-rate classes correspond to the bit-rates of SDH and OTN interfaces [10].

2.2 Constraints

The Petaweb design presents Capacity and Integrity Constraints.

Edge nodes, and optical links have maximal capacities to respect. The *Capacity Constraints* are modular, the resources can be allocated and incremented only through discrete quantities: the optical link capacity can be increased by a multiple of the capacity of W lambda-channels at a time and should be verified for both directions; the capacity of an EN depends on the number of connected optical fibers.

Furthermore, we have to consider additional constraints to satisfy basic communication system requirements in terms of delay and buffering operations; we call them *Integrity Constraints*:

1. All the TLPs of a CR should be transported on the same optical trunk line;
2. All the time-slots associated with a TLP should be transported on the same optical link;
3. All the TLPs of a CR are to be transported contiguously in the time and in the frequency domains.

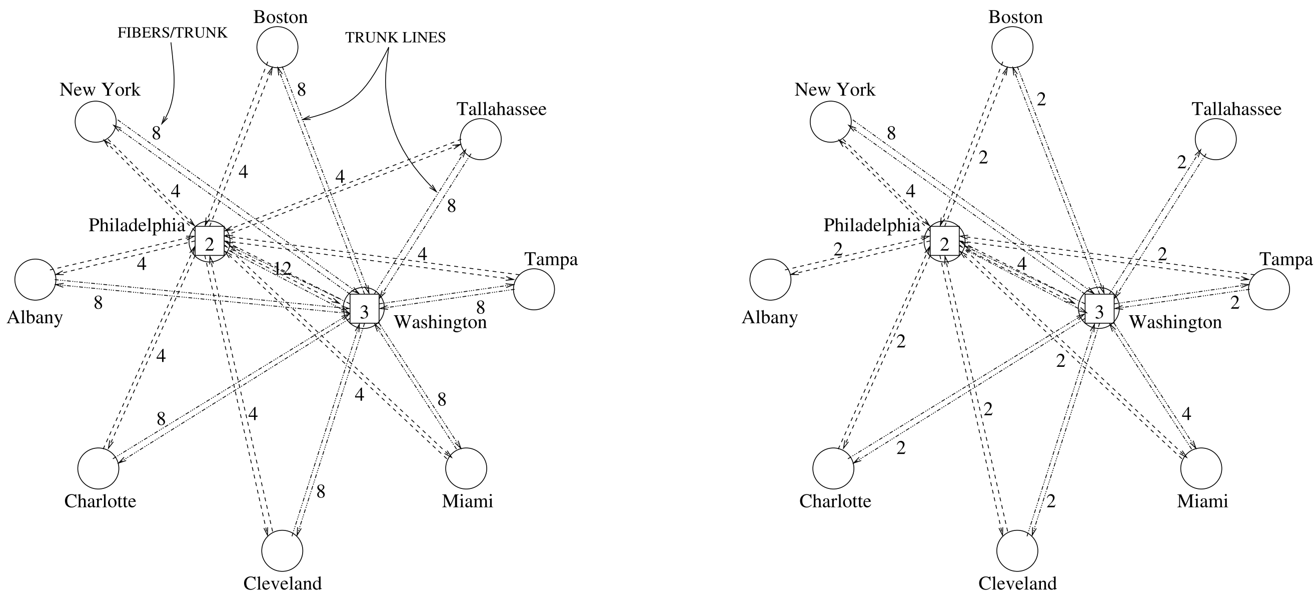
The first Integrity Constraint insures that the traffic between two ENs is switched in the same site. This is like stating that they are transported over the same trunk line. Without this constraint, too much time would be lost at the destination EN because of out-of-order buffering operations: two TLPs of the same connection request may be switched in different sites cumulating different propagation delays.

The second Integrity Constraint insures that the time-slots of a same TLP are switched at the same core node. This is like stating that they are transported over the same optical link (since each edge node is connected with one optical link per core node).

Finally, with the third Integrity Constraint we ease multiplexing/demultiplexing operations, lighten buffering operations at destination ENs, and relate lost data to the minimum possible number of CRs in the case of damage of a single switching plane.

3 Dedicated path protection

Figure 2 reports optimized Petaweb topologies obtained by using the optimization procedure proposed in [6]. The regular topology has all the optical links enabled, and the optical links connected to a CN-r are composed of s_r fibers. Please note that in this particular solution, the core nodes are located at Philadelphia and Washington, but Philadelphia and Washington are also edge nodes. The quasi-regular topology, also proposed in [6], contemplates the deactivation of those fibers that are unused in the optimized regular topology; in this way, the network cost is reduced more than 55% and the network utilization more than duplicates. The quasi-regular topology can be thus determined heuristically by removing unused fibers and ports from an optimal regular Petaweb topology, without any change in the lightpath



(a) Regular topology (10A) (b) Quasi-regular topology (10A)

Fig. 2 Optimized physical topologies without path protection

routes and switching schemes. The drawback is that such a topology is not reliable. To see why, the reader is referred to the Tallahassee and Albany edge nodes in Fig. 2b. They are connected to the transport network through only one trunk line. In the case of failure of one of these trunk lines and if the network operator wants to adopt a restoration strategy, those edge nodes would remain totally isolated from the network and their outgoing and incoming traffic could not be restored at all. Moreover, consider the case where all the core nodes are located in the same switching site (a likely possibility for small networks): all the edge nodes would be connected to the transport network through only one trunk line. It is thus necessary to introduce a protection strategy, which for every working lightpath provides a link-disjoint protection lightpath.

Also notice that even if a switching plane failure affects only the lightpaths switched there, a failure of a whole core node or switching site, e.g. caused by facility flooding, would be a disaster.

We want to design a reliable Petaweb transport network offering restoration functions for its lightpaths and nodes directly in the physical layer. Restoration techniques already exist in the electronic layers (IP, TCP, ATM, SDH), but, even if effective, they require signaling procedures that slow down the restoration process [11]. We need to limit the restoration time because the Petaweb trunk lines may transport optical flows at a Tb/s rate, belonging to various CRs.

Looking for a protection strategy to apply to our network model, we excluded link protection techniques because with the Petaweb architecture the installation of backup trunk lines would require a replica of all the CNs in other switching sites at an enormous cost.

We need to choose a path protection strategy for our network model that does not alter its qualities in terms of simplicity or its switching schemes, and that protects the network in the case of single trunk line failures. The protection functions are performed at the ENs and we want to avoid excessive signaling operations involving the CNs. Because of the high working rate of transmission and switching equipment, any ms of time elapsed for restoration may imply Tbits of data loss. In the event of a trunk line failure, a long signaling phase to establish the protection paths for all the affected TLPs would be required; and it would involve not only the CNs and the EN connected by the faulty line, but even CNs of other switching sites candidate for the TLPs routing. Moreover, the ENs affected by the failure should have knowledge of the actual lightpath topology and routing schemes in order to provide for physical reconfigurations of CNs and compute in some way the optimal protection path. Thus, this signaling phase would be reasonably long, and also hazardous because the CN reconfiguration may not be successful.

We chose a Dedicated Path Protection (DPP) strategy: the protected signal is sent over two separate allocated paths, then the receiver selects one among them. For every working ts-lightpath (wTLP) of our network model we have to allocate a protection ts-lightpath (pTLP). In the 1+1 DPP case the signal is split at the origin over two disjoint paths, and thus there would not be a signaling phase. In the case of 1:1 DPP, the pTLP is sent on the protection path only when the failure occurs, and it makes sense to enable a shorter path for wTLPs, and a longer path to pTLPs to be used in the case of failure along the working one. For this reason, the optimization problem should give priority to wTLPs in the contention for short paths. A shared path protection would for sure guarantee a less expensive network requiring fewer resources, but the signaling would be important and would involve CNs and ENs introducing a significant restoration delay.

In the case of one trunk line failure all the wTLPs must be recovered from the allocated pTLPs. The DPP strategy requires the use of a *protection constraint*: every pTLP must be multiplexed on trunk lines different than those of the corresponding wTLP; in the Petaweb architecture this means that a pTLP must be switched in a different network site than that of its wTLP.

Note that the application of path protection, with the above site-disjointness constraint, guarantees to the network even *node protection* and *switching site protection*: if a core node fails, the TLPs it routed will propagate the failure to the destination edge nodes, which will recover the traffic from the corresponding protection paths; similarly, if a whole switching site is damaged and disconnected in the case of disasters, all the traffic its core nodes switched will be recovered and routed elsewhere.

Therefore, even if the protection constraint defined above is supposed to increase the network cost and the equipment to be deployed, it offers in this specific two-hop optical architecture not only path protection, but also node and site protection. The application of node and site protection constraints in classical mesh networks create irregularities that generate or worsen the bottleneck in some spare links. For the Petaweb, this does not happen.

4 The design problem

The Petaweb design problem with dedicated path protection consists in finding the best composite-star physical topology for the given set of TLPs, respecting the network model (Capacity, Integrity and DPP constraints) and the peculiar composite-star architecture.

The optimization engine takes as input the optimal TLP set for the assigned traffic matrix. Then the physical topology is dimensioned in order to route the TLPs. In the Petaweb a TLP route is characterized by the source node, the destination node and the switching core node. The design dimensioning has to select the CN for each TLP, together with the CN size and location. Then the allocated transport and switching resources, i.e., fibers, wavelengths and time slots, are to be assigned to the TLPs. We need to allocate the resources for routes and fibers to transport the requested traffic volume. The set of possible routes is not enumerated a priori. The virtual TLP topology drives the choice for the physical routes and affects the dimensioning of the physical network. Hence the Petaweb network design is jointly a dimensioning and an assignment problem. It is divided into two sub-problems: the Route and Fiber Allocation (RFA) problem, dealing with the physical resource allocation, and the Wavelength and Time-slot Assignment (WTA) problem, dealing with resource assignment. This last is transparent to the adoption of a DPP protection strategy. Indeed, working and protection ts-lightpaths are already allocated to disjoint optical links by the RFA optimization. For this reason, we do not detail the WTA algorithm presented in [6].

4.1 RFA resolution

The RFA problem consists in finding the optimal location of network elements in order to efficiently switch all the TLPs of the virtual topology; every TLP has to be assigned to its switching CN so that its route and optical links are selected.

This problem can be reduced from the Facility/Plant Location Problem [12, 13]: a set of clients is given and every client has a specific demand for a product; the problem is to optimally locate the plants to send products to the clients, minimizing the global cost expressed by the plant fixed cost and by the transport cost; the plant potential sites are known and the client demand is satisfied by a single plant. For the Petaweb design problem, core nodes are like plants, edge nodes are like customers, the propagation delay cost is like a transport cost. However, while in the plant location problem the product is transported from a plant to a client, in the RFA problem the product uses the plant as transit point. Because of the capacity constraints on optical links, edge nodes and core nodes, the RFA also presents some similarities to the Capacitated Facility Location Problem [14], extension of the plant location problem.

The mathematical notations used in the paper are summarized in Table 1. In particular, δ is a special parameter that is introduced to scale the propagation delay cost of pTLPs in order to assign them longer paths than the corresponding wTLPs path; choosing $\delta < 1$ one forces to route wTLPs on the best paths and pTLP on alternative longer paths (and thus the more δ tends to 0, the longer routes for the pTLPs). We consider indeed as objective function the minimization of the core node cost, fiber cost and propagation delay cost jointly.

The minimization of the propagation delay cost in the objective function not only allows the assignment of the best paths to the wTLPs as explained above, but also the minimization of the fiber distance and thus the outage risk. Indeed, network outages are a function of the overall fiberdistance; for instance, in [15] a provider reports that it experienced 2 to 3 fiber cuts per 1000 miles per year.

Table 1 Notation and abbreviations

CR	Connection request
CN	Core node
EN	Edge node
N_{en}	Number of ENs
TLP	Time slot lightpath (ts-lightpath)
H	Set of TLP classes
TLP-h	TLP of class h
wTLP	Working TLP
pTLP	Protection TLP
M	Set of sites
T	Set of pairs of sites $(M \times M)$, $p \in T$ represents a CR
O_j	Subset of T with pairs of fixed origin site j
D_k	Subset of T with pairs of fixed destination site k
V	Set of core node types
W	Number of wavelengths per fiber
C_{ch}	Capacity of a wavelength channel
CN-r	Core node of type r
s_r	Number of switching planes for a core node of type r
E_r	Number of CN-r specimens that can be enabled in a site
f_r	Fix cost for a core node of type r
K_r	Global cost for a core node of type r
P	Port unitary cost
β	Propagation delay unitary cost
$F_{i,r}$	Total fiber installation cost due to the placement of a CN-r at site i
$c(i, j, r)$	Fiber cost to connect site j to a CN-r in site i
γ	Scaling factor for the port cost
(i, r, e)	Triple representing a CN specimen, $i \in M$, $r \in V$, $1 \le e \le E_r$
C_j	Capacity, in Gb/s, of the edge node in site j, $j \in M$
L_h	Maximal number of TLP-h specimens for a CR, $h \in H$
Z_h	Transport capacity of a TLP of class h
(p, h, l)	Triple representing a TLP specimen, $p \in T$, $h \in H$, $1 \le l \le L_h$
d_{ip}	Distance from the source j to the destination k of p passing by the site i: $d_{ip} = \Delta_{ij} + \Delta_{ik}$
$(p, h, l + L_h)$	Triple identifying uniquely the pTLP of the wTLP (p, h, l)
δ	Scaling factor for the pTLPs' propagation delay cost, $0 \le \delta \le 1$
Ω_w	Set of all the wTLPs, $p \in T$, $h \in H$ and $0 < l \le L_h$
Ω_p	Set of all the pTLPs, $L_h < l \le 2L_h$
Ω	Set of all the TLPs, $0 < l \le 2L_h$
y_{ire}	Indicates if the eth CN-r specimen is enabled in the site i
x_{phl}^{ire}	Indicates if the lth TLP-h specimen of CR p exists and is switched by the CN (i, r, e)
μ_R	Network utilization

The modeling of DPP constraints allows dimensioning a core network with null reprovisioning time, i.e., with null Mean Time To Recovery (MTTR) [16], under single equipment or site failure.

We introduce the following ILP formulation for the RFA problem resolution with dedicated path protection.

$$\begin{aligned} \min G(\overline{y}, \overline{x}) = &\sum_{(i,r,e)} (K_r + F_{i,r}) y_{ire} \\ &+ \sum_{(i,r,e)} \sum_{(p,h,l)\in\Omega_w} \beta d_{ip} Z_h x_{phl}^{ire} \\ &+ \sum_{(i,r,e)} \sum_{(p,h,l)\in\Omega_p} \delta\beta d_{ip} Z_h x_{phl}^{ire} \end{aligned} \tag{1}$$

Subject to:

$$\sum_{r\in V}\sum_{e=1}^{E_r} x_{phl}^{ire} + \sum_{r\in V}\sum_{e=1}^{E_r} x_{phl_p}^{ire} \le 1$$

$$\forall i \in M,\ \forall (p,h,l)\in \Omega_w,\ l_p = l + L_h \quad (2)$$

$$\sum_{(i,r,e)} x_{phl}^{ire} = 1 \quad \forall (p,h,l)\in\Omega \quad (3)$$

$$\sum_{(i,r,e)} C_{ch} W s_r y_{ire} \le C_j \quad \forall j \in M \quad (4)$$

$$\sum_{(p\in O_j,h,l)\in\Omega} Z_h x_{phl}^{ire} \le C_{ch} W s_r y_{ire} \quad \forall j\in M, \forall (i,r,e) \quad (5)$$

$$\sum_{(p\in D_k,h,l)\in\Omega} Z_h x_{phl}^{ire} \le C_{ch} W s_r y_{ire} \quad \forall k\in M, \forall (i,r,e) \quad (6)$$

$$x_{phl}^{ire} \in \{0,1\}, \qquad y_{ire}\in\{0,1\} \quad (7)$$

The objective (1) is to minimize the total network cost. The network cost is linearly computed as sum of the costs of all the enabled core nodes, of the optical fibers between the edge nodes and the enabled core nodes, and of the propagation delays cumulated by the TLPs; the second and the third term are the cost of the propagation delays for, respectively, the wTLPs and the pTLPs. The first Integrity Constraint (Sect. 2.2) is implicitly and partially considered by the propagation delay minimization that drives the TLPs of a same CR to follow the same route.

Equation (2) ensures that the path protection constraint is respected. The pTLP and the wTLP for the same connection request can not be routed on the same trunk line, i.e., can not be switched on the same switching site. Given a switching site and a wTLP, the number of enabled CNs switching that wTLP and the corresponding pTLP in that site must be lower or equal to 1.

Equation (3) expresses the second Integrity Constraint: the traffic transported through a TLP, from an origin EN to a destination EN, must be entirely switched in the same CN and, thus, transported in the same optical links. For every TLP, the number of CNs switching that TLP must be equal to 1.

Equation (4) ensures that the capacity constraint is respected at every EN.

Equations (5) and (6) ensure that the capacity constraint on the optical links between every origin EN and every CN, and between every destination EN and every CN, is respected.

Equation (7) sets the binarity constraint on the variables.

The ILP complexity is hard, but not prohibitive for the assigned instances. By grooming end-to-end TLP-1s and TLP-2s of the same CR in virtual sub-classes consisting of, respectively, 4 time-slots and 4 wavelengths, we could control the number of TLPs, and thus the number of variables and constraints.

4.2 Extraction of a quasi-regular topology

In the dimensioned network only a portion of the transport capacity is to be reasonably assigned. Many optical fibers or links may be totally unused for the following reason: the traffic matrices contain a few peaks of traffic between two sites, and a lot of medium-low values; the peaks will induce high utilization of those optical links connected to the connection request's sites, while the other optical links used for low-rate CRs are to be under-used. Indeed, the activation of a switching plane in the network requires installing one fiber for every edge node.

In order to cope with this inefficiencies, we propose the extraction of a quasi-regular structure from the optimal regular dimensioned network [6]. The quasi-regular structure is to be built as reduction of the composite-star regular one: we pass to the resource assignment WTA algorithm only the fibers really exploitable, the others are ignored. For every optical link, we determine the number of fibers needed by the assigned TLPs, and if this number is inferior to s_r the superfluous fibers are no longer considered. This operation does not require an undifferentiated grooming of TLPs of different optical links over the same fibers. The regularity is preserved and will be attainable with future re-installations of the disabled fibers. After this removal, the final composite-star topology becomes irregular, or, better, quasi-regular. The physical connection between an EN and a CN may become partial, but sufficient for the reserved traffic.

The cost of the quasi-regular structure is expected to be significantly lower than the regular structure cost because the unused fibers are disabled in conjunction to the corresponding CN ports. Obviously, the extraction of the quasi-regular topology does not require changes to the WTA algorithm, because one just changes the fiber number for the optical links. An additional physical hypothesis should, however, be assumed with quasi-regular topologies: the switching planes of a same core node with several switching planes should be able to communicate to each other in order to multiplex lightpaths on the same fiber (if required).

5 Network dimensioning numerical results

In this section we discuss the results obtained implementing the resolution algorithms for the Petaweb design problem. The algorithms have been implemented in C++, and we used the callable library of CPLEX 9.0.1 to solve the ILP formulation of the RFA problem. The tests ran on a CPU AMD Opteron 64 bit 2.4 GHz, 1 MB cache, 16 GB RAM.

Table 2 RFA solution with DPP

MODEL	Regular structure				Quasi-regular structure			
	10A	10B	34A	34B	10A	10B	34A	34B
Objective	4260644	4340130	61419238	77837599	2196002	1829473	27086832	36407731
time	2768 s	2330 s	59.59 h	62.1 h	(same as regular)			
Fiber cost	77.58%	82.25%	82.27%	80.95%	65.85%	71.68%	67.13%	66.33%
CN cost	11.11%	11.80%	5.35%	5.24%	12.21%	14.28%	4.82%	4.15%
Delay cost	11.31%	5.95%	12.37%	13.81%	21.95%	14.04%	28.06%	29.52%
μ_R	23.19%	19.19%	17.38%	13.67%	46.39%	43.18%	46.94%	39.35%

We considered topologies with 10 and 34 edge nodes and with two types of traffic matrices: A matrices contain industrial traffic data, with many zero values; B matrices are, on the other hand, dense and were obtained from a gravity model estimating the traffic between two cities as being directly proportional to the product of the populations and inversely proportional to the square of the distance between the sites. An element of a traffic matrix is a connection request of an origin-destination pair, which is accommodated in the physical topology using one or more TLPs.

The adopted parameter values are: $E_1 = 1$, $E_2 = 1$, $E_3 = 4$ (the number of allowed core nodes of a given type at a site), $\gamma = 0.95$ (the scale factor for the port cost), for 10A and 10B $C_j = 2000$ Gb/s, for 34A $C_j = 4200$ Gb/s, for 34B $C_j = 4800$ Gb/s (the edge node capacities),[1] $L_1 = L_2 = 12$, $L_3 = 20$ (the indexes gap for wTLPs and pTLPs enumerations), $\delta = 0.9$ (the scale factor for pTLPs propagation delay cost).

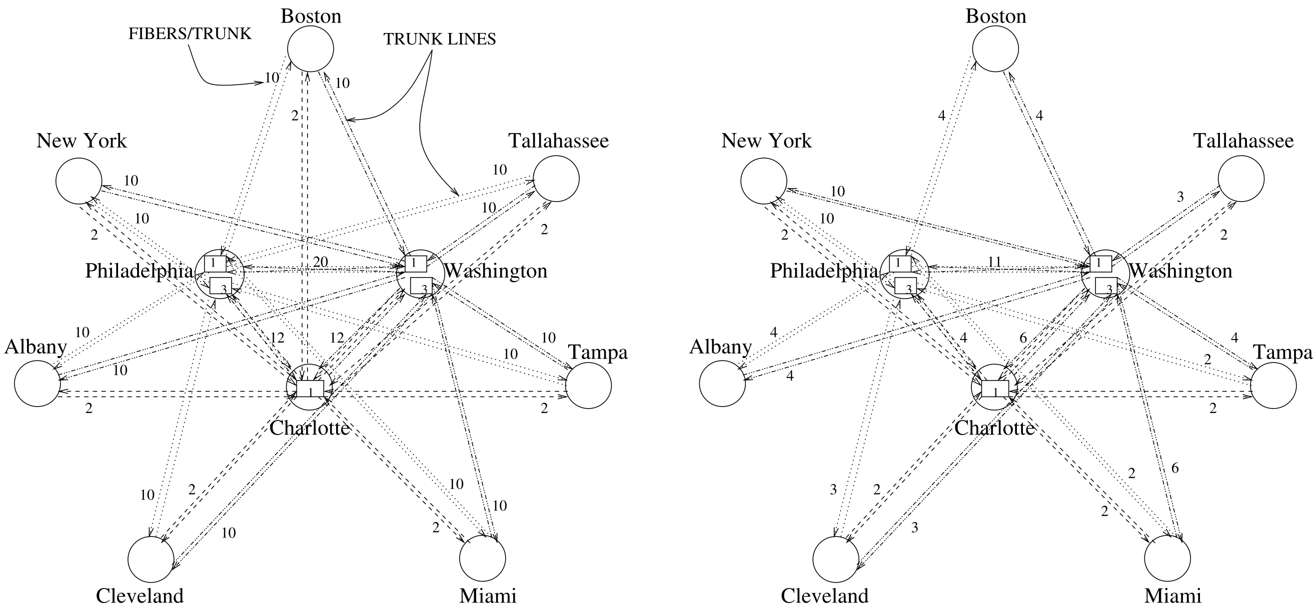
We excluded in the simulations the cases of leased fiber lines in the design dimensioning. For the purposes of performance evaluation, a particular form of the cost function has been used. Since the installation requires s_r fibers per direction and per optical link, we set $F_{i,r} = \sum_j c(i, j, r) = 2\phi(W)Fs_r \sum_j \Delta_{ij}$, where Δ_{ij} is the distance between sites i and j, and F is the cost of a single-wavelength fiber, which is then scaled by a discrete function $\phi(W)$ that depends on the number of wavelengths. We adopted $\phi(W) = W$ considering, thus, that the cost of a fiber is proportional to the number of wavelengths. The other costs are expressed as unit of the fiber cost F: $P/F = 150$ (the port cost), $\beta/F = 0.1$ [km Gb/s]$^{-1}$ (the propagation delay cost), $f_1/F = 20$, $f_2/F = 50$, $f_3/F = 100$ (the core node fix cost for a given type). It is worth noting that the chosen form of $F_{i,r}$ is a theoretical modeling choice, since in the reality the fiber cost is not merely proportional to a unitary cost. As previously mentioned, there are trenching costs, amplifiers, regenerators, cross connect charges, etc. However, when the geographical distances between network sites are very high, and of the same order of magnitude (as for our performance study cases, often hundreds of 100 km), it is acceptable to approximate the end-to-end fiber cost with $F_{i,r}$. In fact, point costs as for amplifiers, regenerators, and edge costs as for fiber trenching and cross connect, can be considered imputing an additional (quite small) constant cost contribution to a length-dependent unitary cost—mathematically, when the distance between sites present a small standard deviation from the average.

[1] Because of the larger amount of traffic requested by the working and the protection lightpaths, we increased E_3 and the ENs capacity constraints C_j with respect to the values used in [6].

5.1 RFA results

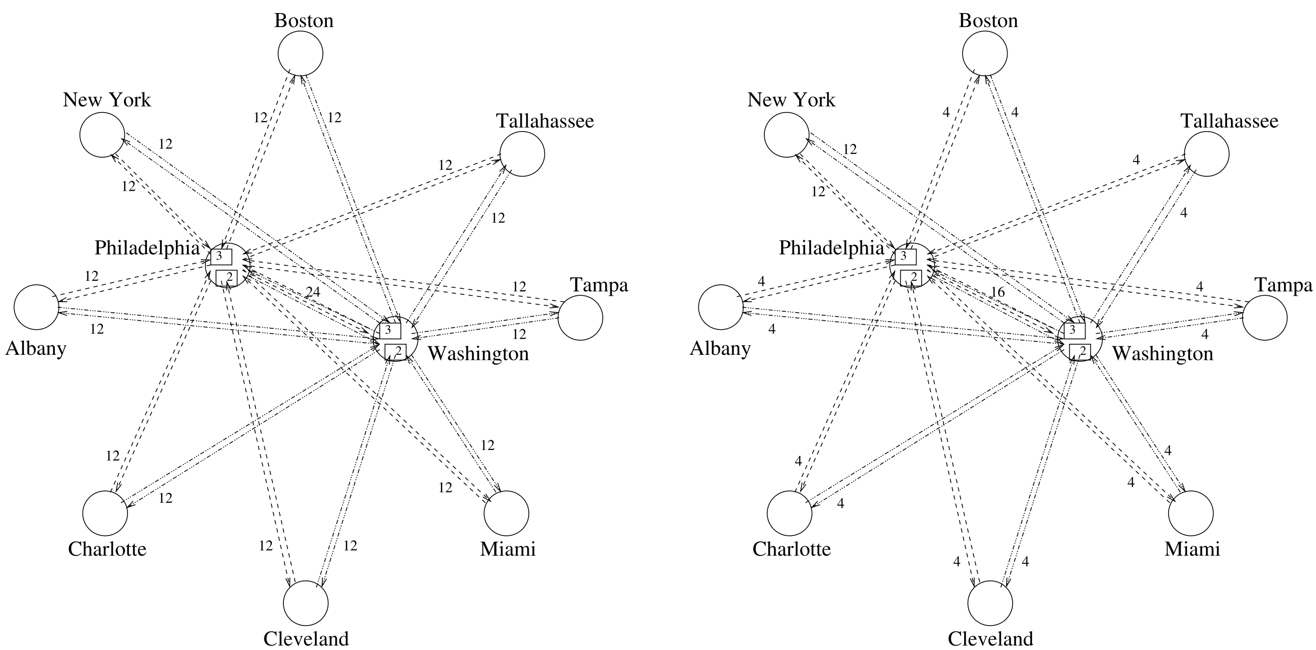
Table 2 shows the results of the resource allocation problem (1)–(7), for the 10A, 10B, 34A and 34B models. The left part refers to the regular topology while the right one refers to the quasi-regular topology. In the tables, μ_R indicates the network utilization, that is, the ratio between the transport capacity allocated for the TLPs and the global allocated capacity, in accordance with the definition given in [17]. Figure 5 displays the CNs geographical distribution for the optimized 34-node networks.

Figures 3 and 4 illustrate the optimized regular and quasi-regular topologies for the 10-node networks. In these figures, circles represent edge nodes and squares represent core nodes. An edge represents a trunk line, and the number over an edge represents the number of fibers per trunk line. Edges connected to the same switching site are equally dashed or pointed lines. As already mentioned, core nodes are not directly connected, but core nodes and edge nodes can be collocated in the same switching site, so that the direct links between Washington and Philadelphia in the figures, for example, represent edge-to-core links. Table 2 shows the results with path protection. In the table, we present the value of the objective, the resolution time, the weight of each of the type of costs in the objective function and the utilization factor for each designed network.



(a) 10A regular (b) 10A quasi-regular

Fig. 3 RFA solution for 10A model with dedicated path protection



(a) 10B regular (b) 10B quasi-regular

Fig. 4 RFA solution for 10B model with dedicated path protection

5.1.1 Network utilization

The results with the A matrices provide better values than with the B matrices for the network utilization (μ_R). The explanation is that the B traffic matrices are dense and present many CRs with low traffic demands; thus, the links used by these CRs are under-used. What would be the changes if the quasi-regular topology can be used? As it can be assessed in Table 4, the network cost is dramatically reduced, around 50%, and the network utilization has more than doubled. This is due to the fact, stated before, that the regular topology demands the allocation of too many unused fibers. The

Fig. 5 Core nodes geographical distribution for 34-node networks

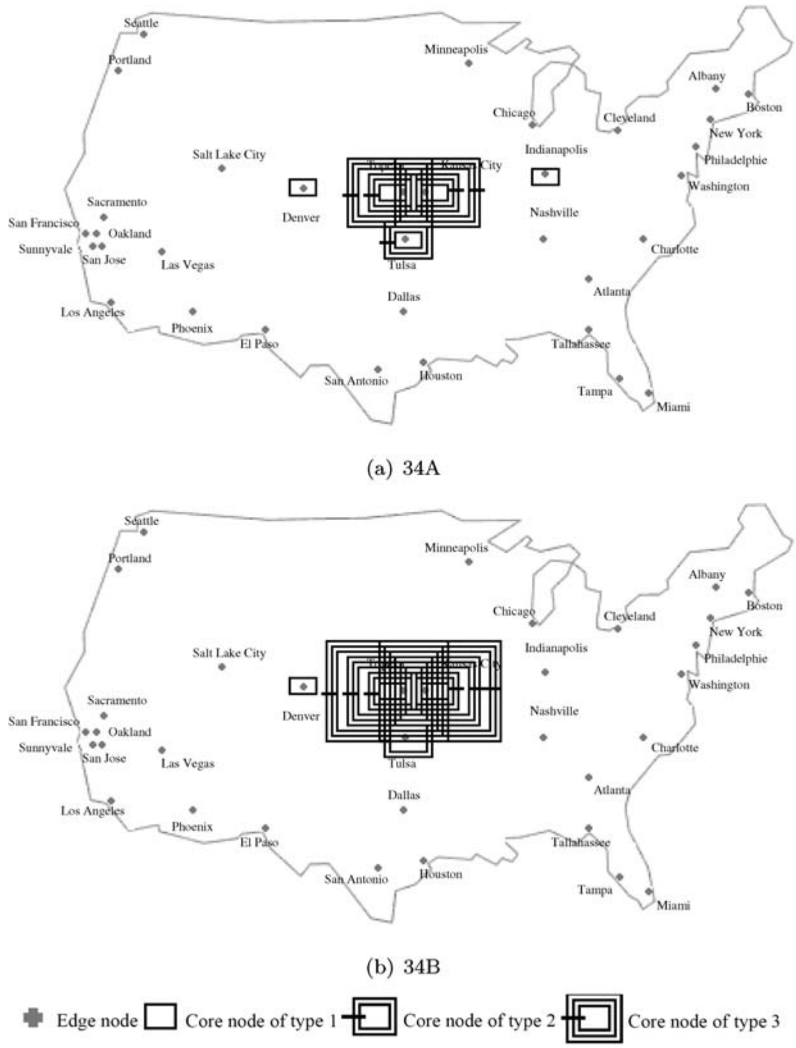


Table 3 RFA solution without DPP

MODEL	Regular structure				Quasi-Regular structure			
	10A	10B	34A	34B	10A	10B	34A	34B
Objective	2281804	2155353	31995440	42596082	982492	840006	12406718	15976542
time	169s	100s	43452s	1685s	(same as regular)			
Fiber cost	77.8%	83.27%	82.46%	81.27%	63.13%	73.65%	63…81%	60.41%
CN cost	11.22%	11.88%	5.44%	5.26%	11.36%	13.88%	4.97%	3.67%
Delay cost	10.98%	4.86%	12.1%	13.47%	25.51%	12.46%	31.21%	35.91%
μ_R	17.91%	15.15%	16.83%	12.39%	48.36%	38.97%	52.69%	51.73%

authors in [4] estimated the quantity of fiber required by a regular Petaweb as roughly 17% larger than that required by a mesh architecture; now, in the quasi-regular topology, the quantity of fiber to install has been considerably reduced; for example, 65% for the case 10A, and 72% for the case 34B. A quasi-regular Petaweb can, thus, compete more concretely with the classical mesh topologies while keeping the advantages of the regular architecture.

5.1.2 *Cost allocation*

Given that the cost reductions are due to unused fibers and disabled ports, the cost allocation changes as well with

Table 4 Increase of the RFA solution with DPP vs without DPP (in % with respect to the case without DPP)

	Regular structure				Quasi-Regular structure			
MODEL	10A	10B	34A	34B	10A	10B	34A	34B
Objective	87%	101%	92%	83%	123%	118%	118%	128%
Fiber cost	86%	99%	91%	82%	133%	111%	129%	150%
CN cost	85%	100%	89%	82%	140%	124%	111%	157%
Delay cost	92%	146%	96%	87%	92%	145%	96%	87%
μ_R	29%	27%	3.2%	10.3%	−4%	10%	−11%	−24%

the quasi-regular topology. The fiber cost weight decreases more than 10 percentage points and, thus, the weight of delay and CN costs increases; the delay cost assumes a weight of more than 10 percentage points than in the case of the regular topology where the fiber cost is over-estimated; the CN cost increases even if its absolute value decreases because the cost reduction due to fibers is more important than that due to port cost. In the case of 34-node networks it is evident that the CN cost becomes unimportant at the expense of the delay propagation cost. And this is even more pronounced with a quasi-regular topology. The presence of high order CNs allows assigning the TLPs to a few trunk lines and to decrease the total number of CNs.

Once again, as it happened in the case without path protection (Fig. 2), the EN in New York fully uses the connected CNs because it has high traffic CRs. The 10A quasi-regular topology presents the trunk lines Charlotte-Boston, Charlotte-Albany and Philadelphia-Tallahassee disabled, but in this case the network remains survivable. And the quasi-regular topology for 10B is fully meshed with a big number of disabled fibers. As a consequence, as expected, with the quasi-regular topology the network utilization increases very significantly, indicating that, on average, the network fibers are used almost at the 50% of their transport capacity. This is a good result considering that, thanks to the time-sharing, we have better exploited the lambda-channels; the idle capacity is available for further network extensions, such as resource re-provisioning or low-level traffic provisioning.

5.1.3 *Changes with respect to the case without path protection*

To analyze the difference in the results with or without DPP, we have portrayed in Table 3 the previous results presented in [6] for the case without DPP and reported the differences in Table 4. The last table presents the increase (in % of the case without DPP) of the objective function, the different costs and the utilization factor. The reader should beware that in Tables 2 and 3 the % for the costs represents the weight of a particular cost (i.e., fiber cost) with respect to the total objective function whereas in Table 4 it represents the percentage increase in cost when DPP is used. Likewise, whereas in Tables 2 and 3 the last row represents the utilization factor, in Table 4 it represents the increase of the utilization factor.

From the three previous tables we can make the following observations:

- The costs increases due to the introduction of DPP are higher for the quasi-regular structure. This was to be expected given that such a topology was chosen for cost reduction to begin with;
- The increase in cost for the regular topology is less than double (except for the 10B case) despite the fact that protection was added to all the connections;
- The fiber cost weight is roughly unchanged for regular networks, while it has increased for quasi-regular networks. However, once again, the increase is not that large;
- The global CNs size has approximately doubled, which can be verified in the result tables that shows an almost double absolute CN cost for the DPP case. Indeed, the CNs disposition is very similar than before as it can be stated that normally the CNs of the case without DPP are to be re-enabled in a dual site to switch the pTLPs/wTLPs of their TLPs;
- In the regular case, the network utilization has increased, the links are thus better exploited than before. On the other hand, for the quasi-regular case, the network utilization decreases for nearly all instances (except the 10B case).

Nevertheless, the execution time has increased, but it is still reasonable. Adding the protection constraint to the heuristic tool described in [3] opportunely adapted to this network model, we could set up good cut-off values for CPLEX.

5.2 WTA results

The WTA task is solved with linear complexity. In this section we report the wavelength and time-slot assignment in

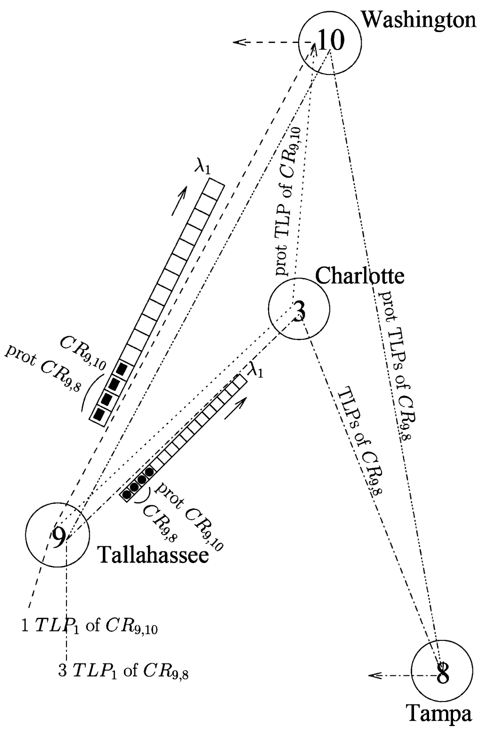


Fig. 6 Routing and assignment in a study case of 10A solution

a case study in order to verify how the pTLPs have been assigned. We analyze, for simplicity, only the optical links exiting the EN in Tallahassee for the 10A quasi-regular solution. As it can be noticed in Fig. 3, using the quasi-regular topology the trunk line Tallahassee-Washington has downgraded from 10 fibers (5 fibers per direction) to 1 fiber from Tallahassee to the CN-1 in Washington and 2 fibers from Washington to Tallahassee. Thus the 4-fibers optical link between the EN in Tallahassee and the CN-3 in Washington has been disabled in the quasi-regular topology. Then, Tallahassee is connected to the CN-1 in Charlotte through one fiber per direction.

The EN in Tallahassee has only two egress CRs, one of 1.64 Gb/s with Tampa ($CR_{9,8}$), and one of 0.2 Gb/s with Washington ($CR_{9,10}$): the first is accommodated using three TLP-1s and the correspondent pTLPs; the second is served by one TLP-1 and its pTLP. The RFA results indicates that on the fiber going from Tallahassee to Washington one must transport the wTLP of $CR_{9,10}$ and the pTLPs of $CR_{9,8}$, and that on the fiber going from Tallahassee to Charlotte one must transport the wTLPs of $CR_{9,8}$ and the pTLP of $CR_{9,10}$. What are the optimal paths for these wTLPs and their correspondent pTLPs? Figure 6 shows the assignment and the routing of the TLPs (only on the egress fibers of Tallahassee). And what is the effect of choosing $\delta < 1$? The path chosen for the 9–8 wTLPs is the shortest one, a total of $858 + 1132 = 1999$ km, while the path for the pTLPs is $1574 + 1799 = 3663$ km.

On both the fibers quitting the edge node in Tallahassee we have 15 wavelengths and 12 time-slots available, yet. This happens because the EN in Tallahassee requests resources for only two CRs with low traffic demand. In a WDM architecture without TDM we would have had two reserved wavelengths, while now only some time-slots of a wavelength are needed. The remaining idle capacity is available to serve further traffic loads, and the two-hop physical connection between edge nodes allows fully usage of this bandwidth.

6 Network robustness performance evaluation

Network robustness refers to several measures related to the reliability, survivability and availability of the designed network with dedicated path protection. Reliability is meant as the capability of a system to durably grant the service, i.e., the capability to reduce end-user perceived failures; survivability represents the capability to rapidly recover a service from a failure that interrupted it; availability signifies the capability to fully access the service when it is required.

We analyze how the Petaweb solution behaves with respect to these aspects with both regular and quasi-regular structures. In our case, a failure can be a trunk line cut, a switching plane damage, a core node disconnection or a switching site disconnection.

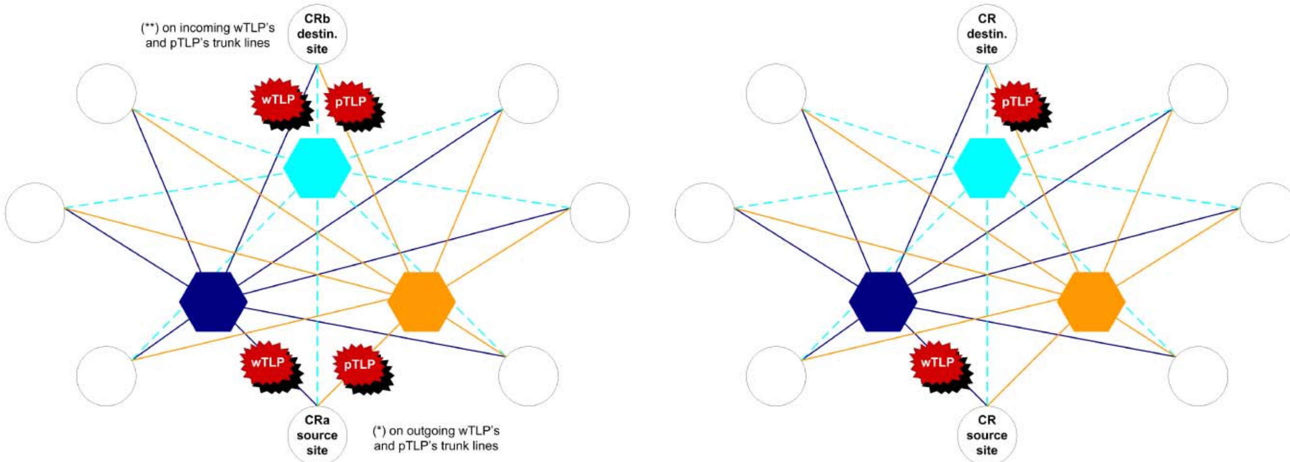
6.1 Reliability and seamlessness

When the network can recover seamlessly, i.e., automatically and immediately, from a network element failure, it can be considered reliable because the failure is not perceived by the end users linked to the edge nodes.

Looking at Figs. 3 and 4 one can notice that the optimized network with the quasi-regular topology is now reliable, while before some ENs could remain isolated after a failure (e.g. see Fig. 2b). Now every EN is connected to at least two switching sites; every wTLP has its trunk-disjoint pTLP where the signal is split. And this stands for trunk line failure, core node or switching site disconnection.

For example, in the case of failure of one of the two trunk lines where a wTLP passes, the destination EN can recover the traffic of the wTLP from the pTLP. Moreover, the case in which all the core nodes are installed in the same site is now not possible: we now have at least two different switching sites, and if a site is totally disconnected, or if a single core node is damaged, all the traffic can be recovered by core nodes in the other(s) enabled switching site(s).

Whether a single switching plane of a core node fails, only the TLPs switched by that plane would be affected. Indeed, even the largest TLP class, TLP-3, requires a single



(a) C.1.1 both egress (*) or both ingress (**)

(b) C.1.2 one egress and one ingress

Fig. 7 C.1: double trunk line failure

switching plane. Whether an entire switching site gets disconnected, or the wTLP or the pTLP of each connection request is also switched in another site, and thus the service is not interrupted.

6.2 Fiber length, protection and failure probability

There exists a trade-off between the level of reliability and the geographical range of the network to be planned. The more the network sites are distant, the higher the trunk line failure probability. The more a trunk line failure is probable, the less reliable the network can be in the case of multiple failures.

Since the Petaweb is expected to offer longer lightpaths than classical mesh architecture solutions, this issue acts as shortcoming for the Petaweb topology, as it will tend to increase link failure probability.

All in all, we can state that the wider the network, the higher the link failure probability. Even if concurrent double failure on disjoint trunk lines—possibly causing an unprotectable outage in a DPP-planned Petaweb—would still remain a very rare event, its occurrence may affect very large volumes. It is thus interesting to assess how much traffic may be reprovisioned in a planned network for such critical failure cases.

6.3 Survivability and reprovisioning

The designed network is thus reliable with respect to single failures of fiber links, switching planes, core nodes or switching sites. It even stays reliable in the case of multiple fiber link, switching plane or core node failures if, respectively, the fiber links are part of the same optical trunk line or the switching planes or the core nodes operate in the same switching site. For such cases in a DPP Petaweb, the reprovisioning time or MTTR [16] is null. Otherwise, one may have service interruption if no other resources are available on alternative paths. In such a case, the interval between the downtime and the next uptime can be an index of the survivability degree of the network.

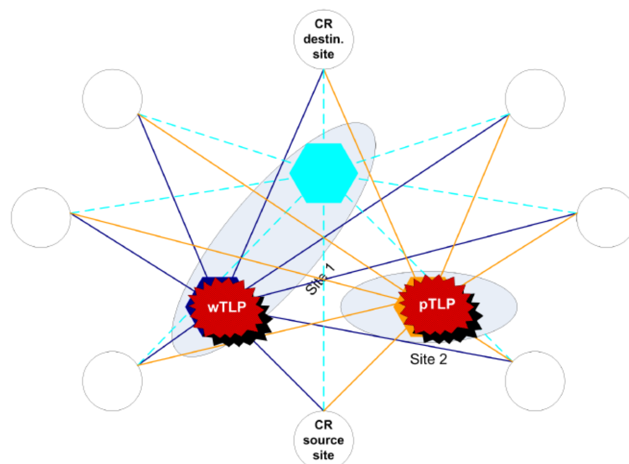


Fig. 8 C.2: double core node failure. For the sake of clarity, a switching site here is represented as composed of two core nodes, while the other of a single one

To recover from multiple failures blocking both the wTLP and the pTLP of a connection request, reprovisioning functions should be implemented at a control plane. Reprovisioning is possible if alternative resources can be instantiated. In such a case, the MTTR is expected to be under the expected MTTR for multi-hop optical networks, since the resource reservation over two optical links can be performed faster than for lightpaths with more than two links and with more than one core switch.

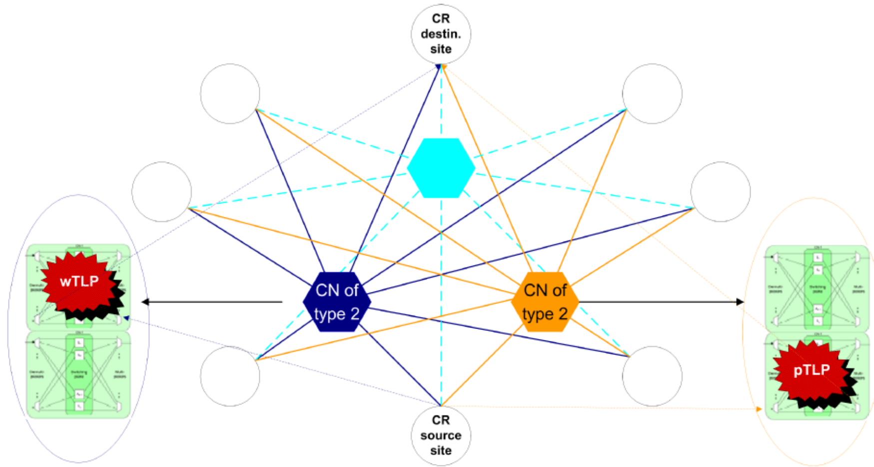


Fig. 9 C.3: double switching plane failure. For the sake of clarity, the core nodes where the failed switching planes are located are represented as being of type 2

Hence, if alternative resources are available, the source edge node should be able to select and instantiate them and reprovide the TLP. To evaluate the survivability of the Petaweb networks with DPP we monitor, for each wTLP-pTLP pair, the network ability to perform TLP reprovisioning in the following blocking cases:

- C.1: double trunk line failure, at wTLP's and pTLP's trunk lines (each TLP using two trunk lines).
  - C.1.1: both the trunk lines are egress lines (Fig. 7a*), i.e. from the source edge node toward the switching core node, or ingress lines (Fig. 7a**), i.e. from the switching core node toward the destination edge node.
  - C.1.2: one failed trunk line is an ingress line and the other an egress line (Fig. 7b).
- C.2: double core node failure, at the wTLP's core node and at the pTLP's core node (Fig. 8).
- C.3: double switching plane failure (Fig. 9), at the wTLP's switching plane and at the pTLP's switching plane.
- C.4: double switching site failure (Fig. 10), at the wTLP's switching site and at the pTLP's switching site.

In what follows, we deeper analyze the results with DPP for all the topology cases and for both regular and quasi-regular structures. For each network case (10A, 10B, 34A, 34B with regular or quasi-regular structure) we look after the chance of reprovisioning a single working TLP when both the corresponding wTLP and pTLP have been interrupted because of multiple equipment failures. We consider the C.1–C.4 multiple failure cases given above.

Table 5 indicates for each network case the ratio of TLPs that could be reprovisioned, having all the wTLP-pTLP pairs been considered under each multiple failure case. For those TLPs that could not be reprovisioned, the pedix indicates the ratio of traffic volume that might be reprovisioned fragmenting the TLP in several lower class TLPs. Hence these two transversal parameters offer a good insight on the survivability of the network solution.

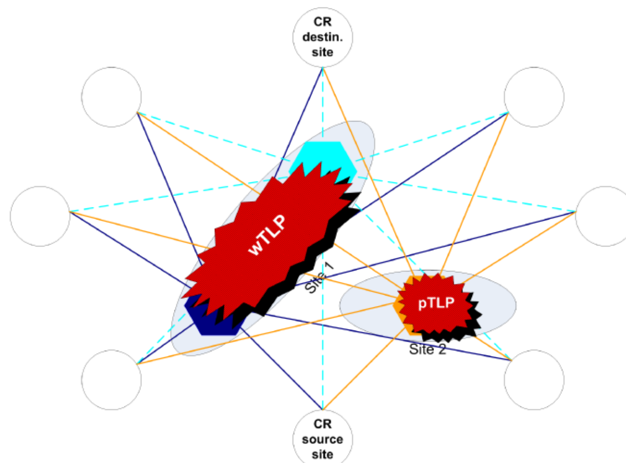


Fig. 10 C.4: double switching site failure. For the sake of clarity, the switching site is represented here as composed of two core nodes, while the other of a single one

From the results in Table 5, we can make the following observations:

- *General observations*
  - Large networks seem to perform better than smaller ones, as all the instances in a 10-node network provided worse robustness results than the equivalent in a 34-node network. This can be explained by the fact that larger switching site diversity and larger resource availability guarantees better TLPs reprovisioning, and this is the case for larger networks.
  - Better robustness results are generally obtained when the traffic matrix is sparse (for the A cases). This can also be explained by larger resource availability produced by the extra sites that are to be considered in the case of the networks designed with the sparser matrix. For example, in Figs. 3, 4 and 5) we have, respectively, 3, 2, 7 (Fig. 5a) and 8 (Fig. 5b) switching sites; indeed,

Table 5 Reprovisioning capabilities under multiple network equipment failures

MODEL	10A		10B		34		34B	
Failure case	reg	q-reg	reg	q-reg	reg	q-reg	reg	q-reg
C.1: double trunk line								
C.1.1: both ingress/egress	$.13_{.40}$	$.7_{.19}$	$.8_{.33}$	$.3_{.15}$	$.57_{.74}$	$.28_{.59}$	$.44_{.68}$	$.25_{.52}$
C.1.2: otherwise	$.25_{.49}$	$.21_{.31}$	$.22_{.41}$	$.18_{.34}$	$.65_{.81}$	$.39_{.72}$	$.66_{.9}$	$.35_{.71}$
C.2: double core node	$.71_{.21}$	$.46_{.15}$	$.63_{.16}$	$.34_{.08}$	$.96_{.23}$	$.74_{.13}$	$.93_{.11}$	$.51_{.14}$
C.3: double switching plane	$.88_{.52}$	$.53_{.24}$	$..69_{.31}$	$.39_{.11}$	$1_{.0}$	$.75_{.82}$	$.96_{.74}$	$.77_{.62}$
C.4: double switching site	$.05_{.10}$	$.04_{.03}$	$..00_{.00}$	$.00_{.00}$	$.31_{.34}$	$.18_{.20}$	$.25_{.11}$	$.8_{.31}$

the more switching sites, the larger path diversity and resource availability there will be in case of failures.

– As expected, quasi-regular structures suffer much more than regular ones from equipment failure and site disconnection. Indeed, a larger switching and transport resource availability guarantees better TLPs reprovisioning. Therefore the decision on the adoption of a quasi-regular or a regular structure might be pondered by a statistical analysis on the failure probability of the different network elements and on the site disconnection probability.

• *Particular observations*

– with double trunk line failures (C.1) we have different survivability features for the two subcases:

∗ with failures on both ingress or both egress trunk lines (C.1.1), a rapid glance to Figs. 3–4 would suggest that the network should have a very low, close to zero reprovisioning ratio for 10-node networks because most of the edge nodes are connected to the backbone through only two trunk lines: when both trunk lines fail these edge nodes would get disconnected. However, Table 5 reports that the reprovisioning ratio is not close to zero, but between 3% and 14% for 10-node networks, and between 25% and 58% for 34-node networks. Indeed, after a closer look to Figs. 3–4 one may note that those edge nodes connected to the backbone with more than two trunk lines (e.g. Philadelphia and Washington sites for 10B), while being collocated with some core nodes, are those likely to be source or sink of most of the traffic.

∗ when the two failed links are one ingress and one egress at different edge nodes (C.1.2), the reprovisioning performance significantly increases.

The TLPs reprovisioning ratio goes over 60% for 34-node regular networks and over 15% for 10-node regular networks, for example.

For those TLPs that could not be totally reprovisioned, the resource reprovisioning ratio may be, however, satisfactory especially for large networks. Seemingly the TLPs that can be easily reprovisioned are those with a low rate.

– with double core node failure (C.2), a majority of the TLPs can be reprovisioned, but those TLPs that can not be reprovisioned are likely to be those with the highest bit rates.

Indeed, we can notice that the pedix—representing the fraction of traffic of the failed connection that could be reprovisioned—are always values around 20% or less, i.e. only roughly 20% or less of the traffic of those TLPs that could not be totally reprovisioned might be reprovisioned (through lower class TLPs). This seems to be due to the fact that TLPs can be easily reprovisioned and switched by another core node, possibly co-located with the failed one, only if their rate does not exceed the idle capacity on the corresponding fiber links (probably equal to a small fraction of the link capacity).

– in the case of double switching plane failure (C.3), TLP reprovisioning is possible with very good statistics, reaching 100% success with regular 34A networks. This confirm the expectations of the authors in [1] about the high switching core reliability of the Petaweb core architecture.

– in the case of double switching site disconnection (C.4), even if the event presents very low probability of occurrence, 10-node networks are almost totally blocked, and 34-node networks can get seriously damaged.

Indeed, the designed 10-node networks dispose of only 2 (10B) and 3 (10A) switching sites, while 34-node networks only 4 (34B) and 5 (34A) switching sites. For the 10B case, double switching site disconnection blocks all switching simply because we have only two switching sites in the dimensioned network. For the 1OA case, only one can survive. For the 34A and 34B cases, only a few switching planes would get overloaded and would create congestion at the edges.

However, whether the network planner disposes of certain statistics about specific failure site, additional site diversity or site avoidance constraints might be easily added to the design model in order to avoid installing large switching equipment in dangerous sites.

6.4 Availability and network extension

The availability is an important criterion to evaluate the performance of a communication network, and in particular of optical WDM networks [18]. A service is a connection request between edge nodes. A network has enough availability if it can offer the service over an existing network configuration, i.e. if it has enough spare capacities and switching capabilities to serve a specific connection request.

Future connection requests might stem from a connection request extension or from the addition of new edge nodes, i.e., new network sites. We have found that optimized survivable Petaweb networks still present a significant amount of idle capacity, roughly 50% of the capacity resources remain available to accommodate further bandwidth requests.

A connection request extension would operationally consist in one or more new TLPs to be provided. The new TLPs' provisioning may or may not be feasible with respect to the available resources and with respect to DPP or delay constraints. Local equipment addition may be required and the upgrade problem may become complex. We treated such a problem in [19], analyzing both the cases of quasi-regular and regular Petaweb network upgrade, concluding that a regular topology is advisable if the network operator has a good initial budget and if frequent upgrades are foreseen, and that a quasi-regular topology is the best choice in the case of low budget and rare upgrades, especially when the upgrade encompasses edge nodes addition.

7 Summary

The Petaweb architecture is an innovative composite-star architecture that drastically simplifies traffic engineering functions thanks to the high level of resource availability and reliability it can offer.

In this paper we proposed a survivable network model and introduced a methodology to design robust Petaweb architectures with Dedicated Path Protection in the cases of regular and quasi-regular structures. The quasi-regular structure appeared to be much more convenient at the expense of minor resource availability. By extensive simulation on real instances we verified that such a network is highly reliable with respect to *single network equipment failures* such as fiber links, core nodes and switching planes, and also with respect to single switching site disconnection.

We checked survivability features in the very low probable case that *multiple network equipment failures* or multiple switching site disconnections block both the working and the protection lightpaths serving a connection request, in order to study events requiring lightpath reprovisioning. The results indicated that multiple switching site disconnections would represent an important issue especially for small networks, and that in such cases the design method should include statistically-inferred site avoidance policies. Moreover, quasi-regular structures acted worse than regular ones given their lower resource availability. The results also indicated, however, that the Petaweb network is highly survivable with respect to multiple core node failure or switching plane failure. In the case of multiple fiber link failure large networks still had a satisfying level of survivability, while small networks would be seriously affected only in the case in which both the egress or ingress fiber links at an edge node get damaged.

Acknowledgements The authors would like to express their gratitude to A. Pattavina and J.-L. Rougier for having supported the collaboration that produced this work, to M. Beshai and F. Blouin of Nortel Networks for proposing the initial design problem. This work was partially supported by NSERC strategic grant nb. STPG 246/59, and the EU IST Euro-NGI, Euro-FGI and Euro-NF Networks of Excellence.

References

1. Vickers, R., & Beshai, M. E. (2000). Petaweb architecture. In *Networks 2000—toward natural network: 9th international telecommunication network planning symposium*.
2. Beshai, M. E., & Blouin, F. (2000). Courteous routing. In *Networks 2000—toward natural network: 9th international telecommunication network planning symposium*.
3. Reinert, A., Sansò, B., & Secci, S. (2009). Design optimization of the Petaweb architecture. *IEEE/ACM Transactions on Networking*, *17*(1), 332–345.
4. Blouin, F. et al. (2001). A comparison of two optical-core networks. *Journal of Optical Networking*, *1*, 56–65.
5. Blouin, F. et al. (2000). Emulation of a vast adaptative network. In *Networks 2000—toward natural network: 9th international telecommunication network planning symposium*.
6. Secci, S., & Sansò, B. (2006). Design and dimensioning of a novel composite-star WDM network with TDM channel partitioning. In *Proceedings of IEEE third international conference on broadband communications, networks and systems 2006*.
7. Dégila, J., & Sansò, B. (2004). Topological design optimisation of a yottabit-per-second lattice network. *Journal on Selected Areas in Communications*, *22*, 1613–1625.
8. Secci, S., Tornatore, M., & Pattavina, A. (2007). Optimal design for survivable backbones with end-to-end and subpath waveband-ing. *OSA Journal of Optical Networking*, *6*, 1–12.
9. Pattavina, A. et al. (2004). Static WDM network planning with TDM channel partitioning. In *Proceedings of IFIP networking conference*.
10. ITU-T International Communication Union (2003). Interfaces for the optical transport networks (OTN). G......709/Y.1331.
11. Maier, G. et al. (2002). Optical network survivability: protection techniques in the WDM layer. *Photonic Network Communications*, *4*, 251–269.

12. Bilde, O., & Krarup, J. (1977). Sharp lower bounds and efficient algorithms for the simple plant location problem. *Annals of Discrete Mathematics*, *1*, 79–97.
13. Karkazis, J., & Boffey, B. (1981). The multi-commodity facilities location problem. *Journal of the Operational Research Society*, *32*, 803–814.
14. Pirkul, H. (1987). Efficient algorithms for the capacitated concentrator location problem. *Computers and Operations Research*, *14*, 197–208.
15. Feuerstein, R. (2005). Interconnecting the cyberinfrastructure. In *Proceedings of cyber-infrastructure workshop*.
16. Grover, W. D. (2004). *Mesh-based survivable networks*. New York: Prentice Hall.
17. Yegenoglu, F., & Sherk, E. (2005). Network characterization using constraint-based definitions of capacity, utilization, and efficiency. *IEEE Communications Magazine*, *43*(2005), 132–138.
18. Arci, D. et al. (2003). Availability models for protection techniques in WDM networks. In *Proceedings of 4th international workshop on design of reliable communication networks (DRCN)*.
19. Secci, S., & Sansò, B. (2007). Upgrade of a composite-star optical network. In *Lecture Notes on Computer Science*. *Proceedings of IFIP NTMS 2007*. Berlin: Springer.

Stefano Secci was born in Milan, Italy, in 1980. He received the M.Sc. degree in telecommunications engineering from Politecnico di Milano, Italy, in 2005, defending a thesis developed at the Research Center in Decision Analysis (GERAD) of the École Polytechnique de Montréal, Canada, and a dual Ph.D. degree in ICT networking from Politecnico di Milano and Télécom ParisTech, in 2009, with a dissertation on multi-provider collaborative network architectures. He is now a postdoctoral research associate at Télécom ParisTech, France. Previously, he also worked as research associate at Politecnico di Milano, as service engineer at Fastweb Italia and as ICT expert consultant for some private and public organisations. His research activity currently covers inter-domain traffic engineering, routing optimisation, IP and GMPLS network design, as well as operations research, game theory and graph theory. He is the recipient of the NGI 2009 best paper award, and author of 20+ articles in refereed international journals and conferences.

Brunilde Sansò was born in Rome, Italy, in 1960. She received the E.E. degree from the Universidad Simon Bolivar, Caracas, Venezuela, in 1981, the M.S. degree in reliability and the Ph.D. degree in operations research from École Polytechnique de Montréal, Montréal, QC, Canada, in 1985 and 1988, respectively. After Postdoctoral studies at the CRT, University of Montréal, and a Research Fellowship at the GERAD, she joined the Polytechnique faculty in 1992, where she has been a Full Professor since 1997. She is currently with the Department of Electrical Engineering, where she is the Director of the LORLAB, a research laboratory devoted to the performance, reliability, design and optimization of operational planning of telecommunication networks. She is Co-Editor of the books Telecommunications Network Planning (Norwell, MA: Kluwer, 1998) and Performance and Planning Methods for the Next Generation Internet (Springer, 2005). Dr. Sansò is a recipient or co-recipient of several awards and honors, among them, the 2003 DRCN Best Paper Award, the Second Prize in the 2003 CORS Practice Competition, the 1995 IEEE/ASME JRC Best Paper Award, the 1992 NSERC Women Faculty Award, and the 1992 FCAR Young Researcher Award. She is an Associate Editor of Telecommunication Systems and has been technical committee member and program chair for major scientific conferences, reviewer for government agencies, and active industry consultant.