Restorability Enhancement and Service-oriented Protection in Radio Access Networks

Zoe Antoniou
Nokia Research Centre
5 Wayside Road
Burlington, MA, 01803
zoe.antoniou@nokia.com

Wee-Seng Soh
University of Michigan
1301 Beal Avenue
Ann Arbor, MI 48109
weeseng@eecs.umich.edu

Abstract

Current radio access networks are based on tree and star topologies, which have no inherent restorability properties. This paper proposes a heuristic topology enhancement method that adds redundant spans and upgrades existing infrastructure cost-effectively, in order to create partially meshed architectures that can provide the desired level of restorability against single span failures. The algorithm is tested using several different variants of restoration mechanisms. Results show that reasonably good solutions can be achieved in a short time scale. Finally, a service-oriented protection and restoration model is presented. Attributes are defined and used to create differentiated protection classes.

1. Introduction

1.1 Motivation and objectives

In recent years, there has been a rapid increase in wireless network deployment and mobile device market penetration. With increasing dependence on mobile devices, there are likely to be greater demands for wireless networks to offer the same reliability as wired telecommunications and data networks. While there has been considerable research in survivable wired networks, it is only until recently that wireless network survivability issues have received some limited attention [1]–[6]. This is partly because financially strapped wireless carriers are still struggling to recover their infrastructure and license costs. Yet, the need to provide reliable wireless access is becoming a critical issue. For instance, a study by the state of New Jersey indicates that wireless E-911 calls account for 43 percent of all E-911 calls [7].

In every cellular network, there exists a radio access network (RAN), which mainly consists of geographically dispersed BTSs, and network controllers, such as the base station controllers (BSCs) in GSM networks, or the radio network controllers (RNCs) in WCDMA Networks. The transmission facilities (also known as backhaul) in existing RANs and upcoming 3G RANs are typically based on star, tree and chain topologies (Figure 1). These structures are highly susceptible to link or switch failures, which could arise due to hardware/software fault, cable cut, and so on. When a failure occurs, wireless access in one or more cells could be lost. According to [11], problems in the backhaul are the uttermost reliability problems faced by wireless carriers. Previously proposed solutions that attempt to improve the fault tolerance of wireless networks, such as those found in [1]–[4], are capable of surviving from base transceiver station (BTS) failures. However, these proposals typically require higher concentration of BTSs to provide adequate overlapping coverage and could be costly to implement. An alternative approach to improving the fault tolerance of existing RANs is enhancing their connectivity and allocating redundant capacity, so as to enable traffic restoration via alternate paths.

In practice, fault restoration can be implemented in multiple protocol layers such as automatic protection switching in the physical layer, self-healing in the ATM layer, and fast rerouting in the MPLS layer. Usually, fault recovery is attempted first in the lower layers and, if not successful, escalated to higher layers. Currently used recovery techniques typically support best-effort service, restoring the connectivity without any QoS guarantees. Some lower layer mechanisms, such as SDH/SONET, can offer 100% traffic recovery within 50ms, but at the cost of 100% redundancy and limited by a ring topology.

Restoration at lower layers is usually faster. Yet, the need for higher layer recovery arises from the following reasons. Firstly, physical and link layer mechanisms have no visibility into higher layer operations and cannot provide node or traffic class protection. Secondly, lower layer granularity may be too coarse for traffic that is switched using higher layer mechanisms. Thirdly, protocols, such as MPLS, have desirable attributes as compared to the recovery approach of a connectionless network. For instance, a recovery path can be "pinned" to avoid the transient instability of dynamic Shortest Path First routing. Finally, higher layer recovery mechanisms decouple fault protection from dependencies on the physical layer, which may differ between networks.

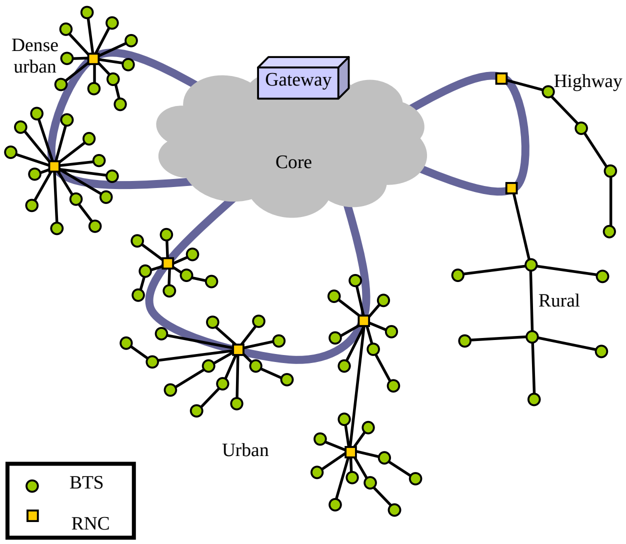

Figure 1: The traditional architecture of the radio access network.

This work presents a protection and restoration method for the RAN. Sections 2 and 3 propose a systematic and cost-effective approach for today's tree and star-like RAN topologies, which have no inherent restorability properties. The proposed approach suggests the incremental addition of redundant spans to the existing topologies in order to create partially meshed architectures. In practice, a mesh restorable network can always achieve 100% restorability when sufficient spares and spans are added to the original network. The challenge is to realize the desired level of restorability while minimizing the upgrading costs. The use of partially meshed architectures to achieve restorability has several advantages as it is described in later sections. Two metrics are defined and used, the restorability and redundancy metrics. A variety of practical issues is discussed reflecting real world constraints, such as modular link capacities, cost model and economy-of-scale effects and split versus single path restoration. The proposed algorithm has numerous enhancements and new contributions as compared to methods found in the literature. Some simulation results are presented in section 4 which demonstrate how the above-mentioned concepts affect the cost-restorability tradeoff and presents the insights gained from these tests.

Given a method for enhancing restorability in the RAN, a service-oriented protection and restoration model is proposed in Section 5. This model aims to meet the protection requirements of applications and users; it uses a set of attributes to define differentiated levels of service and has the following key advantages: (i) integration of protection requirements with the traditional QoS model, (ii) flexibility in the recovery mechanism and traffic granularity, (iii) resource sharing and service differentiation among applications and customer groups, and (iv) implementation of network wide policy-based approaches. This paper does not define how the protection attributes are to be implemented. This is to be explored in future studies. Finally, section 6 concludes and summarises the article.

1.2 Overview of some related approaches

Article [9] presents a Resilience-Differentiated QoS architecture integrating the signaling of the resilience requirements with the traditional QoS signaling of IP services. The service classes in [9] focus on the network-specific QoS requirements, whereas the proposed model considers both network and application-specific needs, consumer distribution and revenue, and addresses the need for RAN topology enhancement.

In [5], a heuristic algorithm for adding new spans to tree-like RAN topologies is presented. Among a set of BTSs that violate the desired loss constraint, a BTS is randomly chosen. A new span is inserted to connect it to one of its neighboring BTSs using a decision rule based on the cost and availability changes that result from the insertion. Span capacities along the backup paths are increased accordingly. The procedure is repeated until all BTSs meet the desired loss constraint. Although the above heuristic approach improves restorability, the cost of the resulting design may be far from optimum. As is shown in [5], the subsequent addition of new spans may render previously allocated capacity unnecessary and no mechanism is proposed to remove such redundancy.

In [4] a two-level hierarchical cell-site architecture is proposed; a macrocell encompasses multiple smaller cells, and acts as a backup system to pick up traffic from any failed cell within its coverage. This requires additional BTSs, and its effect on frequency planning is unclear as it may require the service provider to set aside precious radio spectrum for backup purposes.

In [6], the authors consider the problem of dimensioning a RAN so that it is survivable from single element failure scenarios. Each span's capacity is chosen from a discrete set of capacities. The problem is modeled as a mixed-integer programming problem, and a cutting-plane algorithm combined with heuristics are used to obtain suboptimal solutions. Two alternative models are proposed. The first model uses diversification of the working paths to ensure that a pre-specified fraction of each demand will survive any single element failure without rerouting. The second model considers global rerouting in response to a failure. In contrast to existing fast restoration mechanisms, in which only the disrupted demands are rerouted so as to minimize the impact of a failure, [6] requires all demands to be rerouted after the failure. As a result, the rerouting may be extensive, and could make the network management rather difficult.

2. RESTORABILITY ENHANCEMENT

Tree and star topologies are very sensitive to any type of failure. Since mobile operators have estimated that 30% of their operating costs are associated with the backhaul, any strategy to enhance restorability must be cost-effective. Given the high cost of access capacity and the large number of BTSs in a typical RAN, it can be very uneconomical to use self-healing rings, as they require high-capacity links in every hop. In contrast, partially meshed architectures enable the reuse of existing infrastructure and require lower capacity links. Another distinct advantage is that meshed architectures can provide incremental protection as new spans are added gradually, as opposed to a self-healing ring.

In practice, a mesh restorable network can be fully protected against any single element failure. The challenge is to realize the desired level of protection while minimizing the upgrading costs. This can usually be formulated as an integer-programming (IP) problem, solved by general-purpose solvers. However, the computational requirements are tremendous as the number of constraints and variables increases rapidly with the network size. In addition, it is difficult to obtain data about incremental changes in various parameters. Due to the disadvantages of optimization approaches, heuristics are often used to design restorable networks, although the solutions obtained may be sub-optimal. For instance, consider the two-phase heuristic algorithm proposed in [8] for wireline networks. The first phase (forward synthesis) is a greedy approach that repeatedly adds spare capacity to the network where the greatest increase in restorability will result. The second phase (design tightening) removes any unnecessary redundancy from the design generated by the first phase. It is shown in [8] that this approach could generate near-optimal designs. The algorithm proposed in this paper uses [8] as a starting point and adds numerous enhancements as described in the following subsections.

2.1 Potential Spans and Transmission Capacities

Adding extra spans can enhance the connectivity of an existing RAN. The candidate locations where spans can be added are called *potential spans*. Many existing spare capacity assignment (SCA) algorithms consider only upgrading existing spans to enhance restorability. The proposed algorithm considers both the addition of new spans and the upgrade of existing ones. It needs to select a subset of the potential spans and upgradeable working spans, and specify the corresponding transmission capacities to be used in each span. The selections must be performed in a cost-effective manner, subject to the restorability target.

In a real network, there are many constraints that dictate whether a potential span between a given pair of nodes is feasible, as well as, the set of candidate transmission capacities that may be used in the span. Examples of such constraints include leased line availability, terrain characteristics (e.g., line-of-sight for microwave links), licensing restrictions, distance, antenna space, technological constraints (e.g., interference between microwave channels) and so on. As a result, the set of candidate capacities at each potential span may be different and they are not usually contiguous multiples of a common denomination. In contrast to this work, current SCA algorithms usually assume that all integral multiples of a basic modular unit, e.g. T1/E1, are valid options.

2.2 Modularity awareness

When computing the necessary capacity per span, some existing SCA methods initially perform non-modular capacity assignments, followed by post-modularization of each span's capacity to the smallest module size that can accommodate the non-modular capacity assignment. Since any extra spare capacity resulting from the post-modularization is wasted, this approach results in a solution with higher redundancy than necessary [10]. A better approach is to compute protection capacity in a *modular-aware* fashion, which is adopted in this work.

2.3 General Cost Model

The proposed algorithm can accept a *general cost model* as its objective function and a unique cost can be associated with each span's candidate transmission capacity. The cost can be any user-defined function, such as the true monetary cost to use the transmission facility over a certain period of time. It is important to realize that the cost of the same transmission capacity in different spans may be different. For instance, the cost of a leased line usually depends on its length and availability. For the case of microwave point-to-point links, equipment and licensing costs are affected by the spectrum.

Another advantage of accepting a general cost model is that the algorithm can exploit the presence of any *economy-of-scale* effects in the transmission capacity costs. This is especially important in the case of microwave links, as the capacity costs are expected to be nonlinear, such that a doubling in capacity may only result in a fractional increase in cost. With economy-of-scale effects, it may be more attractive to aggregate restoration paths to form high-capacity spans, rather than spreading the restoration paths over a large number of low-capacity spans. This may sometimes lead to longer restoration paths than those achieved using a linear cost model. The sparsening effect becomes more prominent as the economy-of-scale strengthens. By taking such effects into account, a solution that has a larger total spare capacity but a lower operating cost may still be recognized as a better solution over one that merely minimizes the total spare capacity with no regard for the actual operating cost.

2.4 Restoration Mechanisms

The proposed algorithm supports both *span* and *path restoration*. With *span restoration* (considered in [8]) rerouting occurs between the end nodes of the failed span. Span restoration is fast because the node detecting the failure is, also, responsible for triggering the recovery process. However, it is usually not as capacity-efficient as *path restoration* [12], in which rerouting may occur anywhere between the source and destination.

For span restoration, [8] considers split protection paths, in which a severed span's traffic may be rerouted via multiple restoration paths using any granularity. The proposed algorithm supports the following variants:

- No granularity (NG): affected working path is rerouted via multiple restoration paths.
- Working path granularity (WG): affected working path is rerouted via a single restoration path.
- Entire span granularity (SG): all affected working paths are rerouted via the same restoration path.

For path restoration, there is usually a choice between *failure-independent* (FI) and *failure-dependent* (FD) schemes. In an FI scheme, each working path has one backup path and they are span-disjoint. In an FD scheme, each working path may have more than one backup path to bypass different span failures. In [12], it is determined that the reduction in capacity requirement from using the FD scheme may be insignificant for sparse networks. Hence, in this work only the FI scheme is implemented. The recovery schemes place emphasis on enabling the RAN to survive from single failure scenarios[1].

3. Description of the heuristic algorithm

The proposed algorithm consists of two phases, *forward synthesis* (FS) and *design tightening* (DT). Numerous enhancements and new contributions to the baseline approach in order to accommodate for the practical considerations described in Section 2.

First, two metrics are defined. The first metric is *redundancy*. Let S_i and W_i denote the spare capacity and working traffic requirement of span i respectively. Let I denote the set of all working spans in the RAN. Redundancy is defined as the total working amount of spare capacity in the network divided by the total amount of working traffic:

$$\text{Redundancy} = \frac{\sum_{i \in I} S_i}{\sum_{i \in I} W_i} \qquad \text{(Eq. 3.1)}$$

[1]This assumption enables sharing of spare capacity among the restoration paths for different network elements. If this assumption is violated, though, subsequent failures after the initial failure may not be protected.

The second metric is *restorability*. For span restoration, it is calculated as:

$$R_{\text{span restoration}} = \frac{\sum_{i \in I} R_i}{\sum_{i \in I} W_i} \qquad \text{(Eq. 3.2)}$$

where R_i is the restorable traffic when span i is cut. For path restoration, it is calculated as:

$$R_{\text{path restoration}} = \frac{\sum_{j \in J} R_j}{\sum_{j \in J} W_j} \qquad \text{(Eq. 3.3)}$$

where R_j is the restorable traffic when path j is disrupted, W_j is the working traffic requirement of working path j and J is the set of all working paths in the RAN. So restorability for path restoration is defined as the sum of all restorable traffic among all working paths divided by the sum of all working traffic.

An important utility that both FS and DT utilize repeatedly is a function referred to as *Restorability()*. It is used to compute the RAN's current restorability, using either equation 3.2 or 3.3. It identifies all the feasible restoration paths corresponding to each span failure and it computes the overall network restorability.

Having described the *Restorability()* utility, the FS and DT phases are presented next.

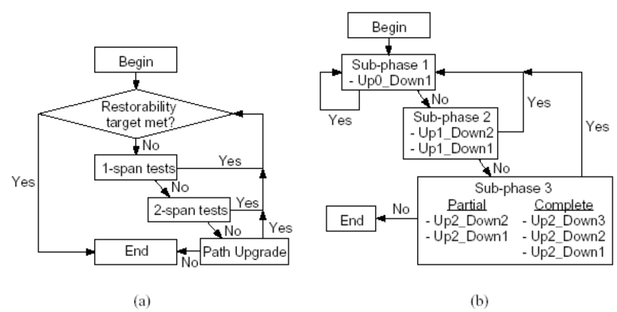

Figure 2: (a) forward synthesis, (b) design tightening.

3.1 Forward Synthesis Phase (FS)

Figure 2(a) shows the flowchart for the FS phase. The *restorability target* is a user input, which must be specified. The FS first checks if the desired restorability target has been met. If not, it enters phase 1, in which a greedy search is performed for a single span within the RAN that can be upgraded to yield the steepest ascend in the restorability vs. cost curve. For each candidate span, the algorithm examines all its feasible capacity choices that are larger than its current capacity, and considers both potential spans and working spans for upgrading. In contrast, the algorithm in [8] only considers adding a fixed capacity size to a candidate span, and uses the best increase in restorability as its selection criterion, with no regard for the actual cost of the upgrade.

If a span is found in phase 1 using the selection criterion, it is permanently upgraded. If the restorability target is not yet met, the algorithm repeats phase 1. If no single span can be upgraded to increase restorability within phase 1, the algorithm enters phase 2, and performs a greedy search for a two-span combination that leads to the steepest ascend in the curve. If an improvement is achieved in phase 2, the two spans are permanently upgraded. The algorithm returns to phase 1 if the restorability target has not been met. If phase 2 does not yield any improvement, the algorithm enters phase 3.

In phase 3, the algorithm randomly picks a span that cannot be restored. Starting from its shortest candidate restoration path, it attempts to upgrade all bottleneck spans along this path. If all bottleneck spans along this path can be upgraded, the algorithm returns to check the restorability target. However, if any of the bottleneck spans is not upgradeable, the algorithm checks the next candidate restoration path and so on until a feasible path is found for upgrading. If no such path is found, the algorithm repeats the procedure for a different span that cannot be restored. If no such path can be found for any of the remaining spans that cannot be restored, the algorithm reaches a stalling point, and exits. When this happens, the RAN has reached its maximum restorability, although it does not meet the restorability target. In order to further improve its restorability, new potential spans and larger capacity links must be included for consideration.

3.2 Design Tightening Phase (DT)

Figure 2(b) shows the flowchart for the DT phase. The objectives of the DT phase are to remove any unnecessary redundancy and to swap expensive combinations of link capacity assignment with less costly ones, while clamping the restorability at the level achieved by the FS phase.

In sub-phase 1, *Up0_Down1*, the algorithm looks at all spans in the network, one at a time, and determines if its capacity can be reduced to the next smaller size without reducing the restorability. Among those spans that satisfy this criterion, the span that yields the largest decrease in cost is reduced in size. If no such span can be found, the algorithm enters sub-phase 2.

In sub-phase 2, the algorithm first attempts *Up1_Down2*. It searches for a 3-span combination in the RAN such that if one span's capacity is upgraded to the next larger size, while the other two spans are downgraded to their next lower size, the cost of the network will decrease by the largest amount without reducing the restorability. Note that the search space of *Up1_Down2* also includes that of *Up1_Down1*. Therefore, if *Up1_Down2* does not yield any decrease in cost, but *Up1_Down1* does, then the latter's solution will be accepted. *Up1_Down2* has higher priority over *Up1_Down1* whenever a solution exists. If either test generates a solution, the algorithm will return to sub-phase 1. Otherwise, it enters sub-phase 3.

In sub-phase 3, there are two options. One option is to run a complete sub-phase 3, which requires a search for a 5-span combination of *Up2_Down3*. This can be time consuming especially for a very large network. Therefore, the algorithm allows the option of partial sub-phase 3, in which only a 4-span combination of *Up2_Down2* is searched. In either case, the larger search space will always contain the smaller search space as before. If no solution is found in sub-phase 3, the algorithm exits and the final capacity assignments have been reached. Otherwise, the algorithm returns to sub-phase 1.

In the original DT phase proposed in [8], only the equivalents of *Up2_Down3*, *Up1_Down2*, and *Up0_Down1* are implemented. Since it regards total spare capacity as the objective, and all capacity options are equally spaced, these are the only combinations that may reduce the objective. However, this paper addresses a more complex problem:

- o different spans have different capacity choices,
- o capacity options may be unequally spaced,
- o capacity costs may exploit economy-of-scale effects
- o capacity costs are location-dependent.

Since the true monetary cost of the network is used as the objective, it is now possible for *Up2_Down1*, *Up2_Down2* , and *Up1_Down1* to reduce the objective.

4. Simulations and Results

4.1 Test Model

An arbitrary RAN topology with two star-like structures is used to test the designed algorithm (Figure 3). Each star consists of 20 BTS, with capacity requirements chosen randomly from the set {2, 4, 6, 8, 10, 12} Mbps. It is assumed that all traffic requirements are symmetrical. The

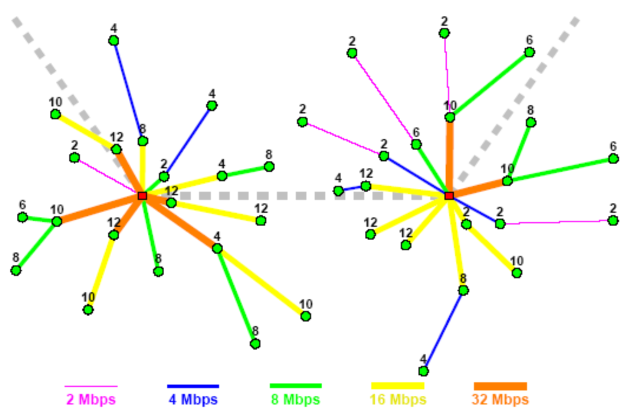

Figure 3: RAN topology for simulation tests.

number above each BTS indicates its capacity requirement. The dotted lines are high capacity links connecting the center of each star to the RNC (not shown). There are 82 locations where potential new spans may be added (not shown), and all working spans are upgradeable if they have not reached their maximum possible capacity.

The following tests aim at restoring the working traffic between the centre of each star and the BTSs for any single working span failure. The objective is to reach 100% restorability while minimizing the cost. For simplicity, no direct inter-BTS traffic is assumed. There is only one possible working path between the centre of each star and each BTS within the star. The initial capacity assigned to a working span is the minimum capacity option that supports the working traffic.

Although the algorithm allows every span to have a different set of capacity choices, for simplicity the same set of capacity values is assumed, namely {2, 4, 8, 16, 32} Mbps. The cost model uses the length of the span, *L*, in addition to a *base cost*, to determine the final span cost. Given a base cost of $C_{base,x}$ for a span with capacity *x* Mbps, the cost of the span is assumed to be:

$$\mathrm{Cost}(\mathrm{L},\mathrm{x}) = C_{\mathrm{base},x} \times \left(1 + \frac{L}{L_{\mathrm{ref}}}\right) \qquad \textbf{(Eq. 4.1)}$$

where L_{ref} is a reference distance. The value of L_{ref} is used to adjust the sensitivity of the cost to changes in the span's length. When L_{ref} is small, an increase in length has a large effect on the span's cost. This could be the case for leased lines, where the cost is normally associated with its length. When L_{ref} is a large number, the cost of the span is less dependent on its length. This relates to the use of microwave links. The base cost of the smallest link capacity (i.e. 2 Mbps) is assumed to be 1 unit for both linear and nonlinear cost models. For a link capacity of x>2 Mbps, the base cost is defined as follows:

$$C_{\mathrm{base},x} = \begin{cases} x/2 & \text{linear cost model} \\ 1.5^{\log_2 x - 1} & \text{non-linear cost model} \end{cases} \qquad \textbf{(Eq. 4.2)}$$

Note that the nonlinear cost model is 4x3x, meaning that as the capacity quadruples, the cost triples. The nonlinear cost model favors the use of existing spans as opposed to the addition of new spans. This cost model is used for the purpose of the simulations in this paper. Any cost model can be incorporated in this approach.

4.2 Simulation Tests

Table 1 summarizes the simulation tests and results. For span restoration, a hop limit of *H=5* was imposed on the restoration paths of each span. For path restoration, *H=7* was imposed on the restoration paths. The restorability target for each test was set to 1.0. In addition, test SG was formulated as a pure IP problem with binary variables. The SG is selected among all the tests because its problem size is the smallest compared to the rest. The formulation was encoded in MPS format, and passed to two general-purpose solvers that utilize branch-and-bound techniques, namely, IBM OSLMSLV and Mosek. They were allowed to run on a 1.8 GHz machine for more than 48 hours each, and the best binary solutions discovered up till the time of manual termination were recorded.

From Tests NG-L and NG-NL, which differ only in the cost model, it is observed that the use of a linear cost model results in larger number of extra spans. A closer look at the network designs (not shown here) reveals that 50% of the extra spans in Test NG-L have very small capacity (2 Mbps), as opposed to 24% in Test NG-NL. This is due to the economy-of-scale effect introduced by the nonlinear cost model. Note that the extra cost of NG-L should not be compared with that of the other tests as they are derived based on a different cost model.

Test name	Test			Results		
	Span/Path restoration	Property	Cost model	No of extra spans	Extra cost (%)	Redundancy (%)
NG-L	Span	No granularity constraint	Linear	48	41.91	87.06
NG-NL	Span	No granularity constraint	Nonlinear	33	62.26	96.52
WG	Span	Granularity of working path	Nonlinear	34	88.65	146.27
SG	Span	Granularity of an entire path	Nonlinear	35	97.06	177.61
FI	Path	Failure-independent restoration	Nonlinear	30	80.08	145.27

Table 1: Summary of simulation tests and results.

Next, the results of Tests NG-NL, WG, and SG are compared. As can be seen, when the granularity of the restoration paths become coarser, the resulting extra cost, number of extra spans and redundancy increase. This is because each restoration path carries more traffic as the granularity becomes coarser, which results in a larger number high capacity spans.

Finally, Test FI shows the result for failure-independent path restoration. Since each affected working path is rerouted only via a single protection path, its granularity is equivalent to that of WG test. Comparison shows that path restoration results in lower cost, as well as, smaller number of extra spans being added.

For Test SG formulated as a binary IP problem, both IBM OSLMSLV and Mosek solvers were unable to obtain any global optimum solution at the end of 48 hours, due to the large problem size. However, each solver returned the best feasible solution found during the allocated time. The IBM OSLMSLV returned an extra cost of 115.89%, while Mosek returned an extra cost of 102.84%. The proposed algorithm obtained a cost of 97.06% in merely 10 seconds.

5. Protection model

Sections 2, 3 and 4 present a flexible restorability method which builds on the idea of mesh restorable networks subject to several practical considerations. This enables the introduction of a service-oriented protection and restoration model.

5.1 Objectives

The service-oriented protection model has the following objectives:

i. Policy-based protection which can potentially result in a uniform protection scheme throughout the network, even in cases where the implementation varies in different domains.
ii. Protection schemes that can guarantee both network and application-specific requirements.
iii. Protection service differentiation which can maximize the use of available resources with respect to efficiency and revenue.

The proposed model offers the integration of protection requirements with the traditional QoS model and provides flexibility in the recovery mechanism and traffic granularity. As part of this model, application-specific requirements, customer distribution and revenue are considered along with network-specific requirements. As a result, it offers more control over the tradeoff of protection versus cost and complexity.

5.2 Recovery process

Although there are numerous implementation-related issues, the skeleton of the recovery process can be summarized as follows. Primary paths are established between the Radio Network Controller (RNC) or RAN Access Gateway (AGW) and each BTS. Each connection contains an aggregate of streams belonging either to a single class of service with a homogeneous service profile, or multiple traffic classes. Protection (secondary) paths are set up based on the service model. Upon failure detection, a failure notification is communicated to the node which initiates the restoration procedure, also called point of repair (PoR). The PoR initiates the restoration procedure based on the protection service model. Periodic re-optimization of the protection paths can be employed based on updated information about the network topology and policy, traffic and user profiles.

The proposed model allows for resource sharing. If the protection paths are computed by a centralized server, the server can compute the restoration paths for each protected network element independently, one at a time. For every new protection path being computed, the bandwidth required for other protection paths that are not activated simultaneously is ignored, thus, implicitly ensuring that the spare capacity can be shared. No explicit bandwidth reservation or signaling is needed for each protection path.

5.3 Attributes

An important aspect of the protection model is differentiated treatment between service classes. This is defined in terms of several attributes presented next.

Protection model attributes	
QoS	Equivalent to QoS prior to failure, limited QoS or best effort
Protection scheme (possible options depend on technology used)	Some options: (a) protection switching: in advance protection path computation end-to-end and resource reservation; requires a notification process between the point of failure detection and the PoR, which can result in delay, (b) MPLS fast rerouting (FRR): protection path is pre-computed and pre-reserved, can cover the primary path either locally or globally; FRR does not require signaling to trigger the recovery process, (c) on demand rerouting: the least conservative option.
Routing precedence	Protection path can be pre-established, pre-computed and established on demand, computed and established on demand or none.
Resource allocation	Pre-reserved, reserved on demand, none.
Resource use	Dedicated, shared or extra low priority, pre-emptible traffic allowed on protection path under normal working conditions.
Pre-emption or retention priority	Protection path for a specific class can be pre-empted or retained as compared to the varying needs of other protection classes.
Recovery scope	Global (end-to-end), local.
Failover time	Can be categorized based on the traffic profile, e.g. 10-100ms, 100ms-1s, 1s-10s or no guaranteed limit.

Table 2: Protection Model Attributes.

In an effort to optimize the use of resources, it is possible to employ *n-to-1* protection (*n* working paths protected by *1* recovery path). If the intent is to protect against any single failure, the *n* working paths should be *node* and *link disjoint* so that there are no contending demands for a recovery path by two or more primary paths from any single failure. This can lead to more efficient resource utilization at the expense of more complex computations. Another option is *1:1* (one for one) recovery paths which can be used to transport low priority pre-emptible traffic under normal working conditions. *Constrained Based Routing* (CBR) can be used in the computation of such routes. It can be performed on-line or off-line and can incorporate any variety of constraints.

5.4 Protection classes

The above-mentioned attributes are used to defined protection classes:

A. Protection traffic classes: associates the protection service model with the traditional QoS model. It classifies traffic based on its QoS requirements and each QoS class is associated with a protection class. For example:

i. RT: equivalent QoS, predefined path, fast failover time, highest retention priority
ii. NRT: equivalent QoS, predefined path, fast failover time, medium retention priority
iii. Background: best effort QoS, path on demand, slower failover time, lowest retention priority
iv. Best effort: dropped

B. Protection service classes: protection classes are different from the QoS classes. Model classifies services and customers into protection classes based on application-specific requirements, customer revenue and distribution. For example:

i. Platinum customers: corporate, large business, campuses with platinum protection service (e.g. equivalent to A(i))
ii. Premium customers: commercial centres, campuses, residential, small businesses with premium protection service (e.g. equivalent to A(ii))
iii. Bronze customers: rural residential and small business with bronze protection service (e.g. equivalent to A(iii)).

6. Conclusions

Traditional RAN topologies are mainly based on tree, chain and star structures, which have no inherent restorability properties. This work introduces a method to upgrade existing RAN topologies cost-effectively so as to improve their restorability against single span failure scenarios. The proposed approach uses the incremental addition of redundant spans, as well as, the upgrade of existing spans, in order to create partially meshed and restorable architectures. The objective is to realize the desired level of restorability while minimizing the upgrading costs. The proposed algorithm has numerous enhancements over existing heuristic approaches. A number of practical issues are taken into consideration, such as potential spans and candidate transmission capacities, modularity awareness, a general cost model and different variants of restoration mechanisms.

A service-oriented protection and restoration model is, also, introduced. The model can facilitate policy-based protection, schemes that can guarantee network, as well as, application-specific requirements and protection service differentiation to meet requirements and maximize the use of available resources. It employs several attributes to create service differentiation. This offers control over the tradeoff of protection quality versus cost and complexity.

From the tests on an arbitrary RAN topology, observations are summarized as follows. The presence of economy-of-scale effects in capacity costs results in a smaller number of extra spans being used. Cost and redundancy requirements increase when granularity of restoration paths becomes coarser. Also, path restoration results in lower cost and smaller number of extra spans.

7. References

[1] U. Varshney and A.D. Malloy, “An integrated approach for improving the dependability of the emerging wireless networks,” in *Proc. IEEE GLOBECOM*, vol. 6, pp. 3693–3697, 2001.

[2] T.A. Dahlberg and J. Jung, “Survivable load sharing protocols: a simulation study,” *ACM/Baltzer Science Wireless Networks Journal (WINET)*, vol. 7, issue 3, May 2001, pp. 283–296.

[3] Y.H. Wang, W.S. Soh, M.Y. Tsai and H.S. Kim, “Survivable wireless ATM network architecture,” in *Proc. IEEE IC3N*, Oct. 2000, pp. 368–373.

[4] T. Dahlberg, S. Ramaswamy, D. Tipper, “Issues in the survivability of wireless networks,” in *Proc. IEEE MWCN*, May 1997.

[5] A. Szlovencsák, I. Gódor, J. Harmatos, and T. Cinkler, “Planning reliable UMTS terrestrial access networks,” *IEEE Communications Magazine*, Jan. 2002, pp. 66–72.

[6] D. Alevras, M. Grötschel, P. Jonas, U. Paul, and R. Wessäly, “Survivable mobile phone network architectures: models and solution methods,” *IEEE Commun. Magazine*, March 1998, pp. 88–93.

[7] State of NJ, “Report on the New Jersey wireless enhanced 911 system trial: the first 100 days,” June, 1997.

[8] W.D. Grover, T.D. Bilodeau, and B.D. Venables, “Near optimal spare capacity planning in a mesh restorable network,” in *Proc. IEEE GLOBECOM 1991*, pp. 2007-2012.

[9] A. Autenrieth *et al.*, “Engineering end-to-end IP resilience using resilience-differentiated QoS”, *IEEE Communications Magazine*, Jan. 2002.

[10] J. Douchette and W.D. Grover, “Influence of modularity and economy-of-scale effects on design of mesh-restorable DWDM networks,” *IEEE Journal in Selected Areas in Communication.*, vol. 18, pp. 1912–1923, Oct. 2000.

[11] C. Harler, “The reliability of wireless networks-how much safety can you afford?”, America’s Network, January 1998.

[12] Y. Xiong and L. Mason, “On state-independent and state-dependent path restoration in self-healing networks”, Proceedings of IEEE ICC, June 1998, pp. 1114-1118.