Network state advertisement and $p$-cycle protection for reliable connections

PhD Thesis Booklet
2011

János Szigeti
szigeti@mit.bme.hu
Supervisor:
Tibor Cinkler, PhD
1 Introduction

Nowadays the core and metro networks have to provide ever increasing bandwidth for information transmission. These networks are based mainly on optics. The communication between the end nodes in these networks is established over predefined end-to-end connections, i.e., provisioned channels. The provisioned channels have to satisfy predefined Quality of Service (QoS)\cite{1}\cite{2} constraints. To achieve this objective, the QoS parameters are to be taken into consideration while provisioning a communication channel, i.e., when selecting and reserving network resources.

Networks that use optical devices are traditionally multilayer networks, e.g., they have IP/MPLS in the uppermost layer and aSDH/OTN beneath. In current network architectures the different layers are usually controlled separately, and the upper layers may use the services provided by lower layers. This control model is referred to as Overlay model. There is another control model, called Peer model, which allows bidirectional interaction between the neighbor layers. This solution results in a more flexible and more efficient network resource usage. The price of this flexibility is the complexity of the implementation. Beside of the Overlay and Peer models, a so called Vertically Integrated multilayer Control Plane (CP) model has been defined. This model has the principle of having a single, integrated control plane for all the different layers. The vertically integrated CP model offers the most flexible way of network control.

Regardless of the choice of the control model, whenever it comes to the subject of connection provisioning in the optical layer, we have to face the problem of wavelength assignment (as part of the Routing and Wavelength Assignment, RWA). The origin of this problem is the fact that, compared to conventional MPLS routing, in the optical domain the route of a connection is not defined merely by the list of consecutive network nodes. Also a free wavelength channel must be found between each adjacent node-pair along the connection. Moreover, the switching from the ingress to the egress wavelength channel must be resolved at each node. This problem can be easily solved in two extreme scenarios, either when all the network nodes support full wavelength conversion (referred to as Wavelength Interchangeable Network), or when none of the nodes is capable of wavelength conversion (Wavelength Selective Network). Otherwise the wavelength assignment is a complex task.

Different switching devices have different switching and wavelength conversion capabilities. It is the operator of the network who decides what devices to deploy in the network. Thus there can be evolved heterogeneous networks with switching devices of different type and capability. In such heterogeneous network environment the conventional link state advertisement protocols face a serious challenge, since as the number of wavelength channels
keeps growing, the number of control messages to signal connection setup or release or to describe state changes also grows.

In case of vertically integrated control plane, the problem is even more complex, as the attributes and constraints of the different layers have to be taken into consideration simultaneously.

Reliability and availability of network connections can be improved, if we make ready the connections for network device failures and therefore we assign backup (protection) paths to the default (working) communication paths. There are several protection and resilience schemes to realize this resilience [4]. However, how to apply such resilience schemes is a hard challenge both in multilayer and in multidomain networks.

2 Research Objectives

Routing algorithms primarily focus on finding a set of paths between node pairs that fulfill predefined QoS requirement. Another objective of the routing algorithms is the resource efficiency. It is clear, that the performance of the routing algorithm depends on multiple factors, however, it can be stated generally, that one of the most important preconditions of routing is that the Path Computation Element (PCE) [5] needs to have an accurate view of the state of the network and of the network devices, especially in case of routing the requests dynamically. Furthermore, it has to be considered, that the routing algorithm has to be feasible and secure. Finally, when examining the performance of the routing algorithm, it is important to know, how much can we rely on the evaluated or estimated performance metrics.

Accordingly, in my work I have examined the following three problems:

• Regarding the state advertisement, how can we provide an accurate, still compact view of the network state?

• How can we apply p-cycle protection [6] in a multidomain environment?

• What is the accuracy of a commonly used Serial-Parallel availability estimation heuristic [7]?

Examining the first problem, in intra-domain context, we expect to have accurate view of the actual network state and we try to describe this accurate information as compact as possible.

In case of inter-domain routing, resilience is of high importance. When provisioning resilient inter-domain connections we have to take into account, that different resources of
different domains may share a common risk group (Shared Risk Link Group, SRLG [8]). This implies that the classic “single failure assumption” cannot be applied in such cases, thus we cannot rely on conventional end-to-end protection schemes. Due to this fact, other protection schemes, employing pre-defined backup paths, e.g., \( p \)-cycle protection, come into view. Within the topic of inter-domain routing my research has focused on the realization and analysis of the \( p \)-cycle protection scheme applied to inter-domain connections.

There are several ways to estimate the availability of network connections. To derive the connection availability in a network with a complex protection scheme (e.g., shared protection), there are heuristics, which provide a fast, but inaccurate approximation. In my research I have examined the availability estimation of \( p \)-cycle protection. First, I tried to answer, how to get the accurate estimation on the availability of \( p \)-cycle protected connections, furthermore, I have analyzed the approximation error of the Serial-Parallel (SP) availability estimation heuristic [7].

3 Research Methodology

Regarding the network state advertisement, the most important question is the modeling. We have to find a general, versatile model that can be used in diverse scenarios. Nevertheless, the model should also remain as simple as possible. For simplification we have used decomposition: First, we have structured the problem space, decomposed it into pieces and traced back the compound parts to basic problems. Second, the whole model has been constructed of modules which can be combined with each other. The resulting model has been analyzed: Performance metrics, such as resource requirement, required storage space, amount of transferred information and required calculation steps were derived.

We have often used results of graph theory in our protection-related research. Our research related to inter-domain protection relies on methods of graph-transformation.

Examining the problem of availability estimation requires basic knowledge within the fields of probability theory. Furthermore analytic methods were used here.

Simulations were also carried out to study and validate the performance of the proposed methods. As a side-effect of our research, many simulation tools have been developed. These simulators I developed are used at the Department even today. The most important tools are \texttt{idr} (Inter-domain Router) and its successor \texttt{arr} (Advanced Resilience Router), to simulate routing and state advertisement in multidomain multilayer networks with different protection schemes; \texttt{pcycle}, which simulates inter-domain \( p \)-cycle protection; \texttt{aszim}, which is an event driven simulator to monitor traffic on control channels of different controlling solutions. Theoretical results shown in the next sections were supported by simulation
4 Results

As already mentioned, our research area is quite widespread. The research results are organized into three parts: First, in Sect. 4.1, I present the theses related to the dynamic link state advertisement. Next, in Sect. 4.2, we propose and evaluate a protection scheme that is based on the $p$-cycle protection [10] and can be applied to inter-domain connections. Finally, in Sect. 4.3, results related to $p$-cycles (candidate cycle search and availability estimation) are presented.

4.1 Dynamic Topology Information Advertisement to Support Routing

In optical networks, as the wavelength division multiplexing gets denser and denser, the amount of advertised topology information leads to scalability problem. The problem becomes even more complex if we want to apply different protection schemes in an efficient way as the protection schemes require additional information to be advertised (e.g., shareability information). We face another problem if we want to apply different protection schemes in the same network simultaneously. Thesis Group 1 proposes solutions on these scalability problems.

To provide a high level of flexibility (allowing vertically integrated control, traffic grooming, path fragmentation and de-fragmentation or resource sharing) for the optical network, we model the network as a wavelength graph (WG) [C10, C11, C12, C13, C14, C15, C16, C17, C18]. However, due to the specialties of optical switching, the WG is redundant. The conventional Link State Advertisement (LSA) solution assigns an LSA message for each edge of the WG. This way, the conventional LSA inherits the redundancy and the redundant behavior of the WG. I have elaborated inra-domain link state advertisement methods that exploit this redundancy and, compared to conventional Link State Advertisement (LSA) methods, result in a reduced amount of transmitted information.

4.1.1 Rule-Based Topology Advertisement (RBTA)

The state changes in a WG due to resource allocation or deallocation are correlated, thus conventional LSA methods carry redundant information.

**THESIS 1.1 ([C1] [C2]):** I have proposed the RBTA (Rule Based Topology Advertisement) method for lossless and compact topology information advertisement. In case
of dynamic connection provisioning, the amount of information advertised by RBTA is independent of the size and complexity (wavelength and port number) of the network devices, thus, in heterogeneous optical networks, the RBTA outperforms conventional LSA methods.

The basic idea of the RBTA model is to formulate this redundancy by means of rules. As the redundancy comes from the structure of the optical crossconnects, during the operation phase of the network these rules – just like the structure of the crossconnects – remain invariant. Hence, it is sufficient to send that rule-set to the PCE(s) only as an initialization, and afterward, whenever a state change happens that affects multiple WG edges, only the triggering cause of the state change needs to be advertised. Knowing the rules and the cause, the PCE is able to determine all the evolving link state changes in the WG.

To define the rules we need the following information for each WG node:

id identifies the port.

TE_attr is the compound of various TE attributes (e.g., total_cap denoting the maximal capacity of the port).

in_links is the set of ingress link identifiers.

in_max is the maximal number of links simultaneously switched at the ingress of the port.

out_links is the set of egress link identifiers.

out_max is the maximal number of links simultaneously switched at the egress of the port.

Using these notations the switching rule for each port is given as a compound of data:

\[ \text{Rule}(id): (id, \text{TE}_\text{attr}, \text{in}_\text{links}, \text{in}_\text{max}, \text{out}_\text{links}, \text{out}_\text{max}) \]

Being aware of this rule set, the advertised information about the \(i^\text{th}\) state change can be encapsulated into a conventional LSA message, as it should only contain that an edge \(\text{link}_\text{id} \in \text{in}_\text{links} \cup \text{out}_\text{links}\) suffered a given amount \(\Delta\text{cap}\) of capacity change:

\[ \text{LSA}(i): (\text{link}_\text{id}, \Delta\text{cap}) \]

As we assign this fixed sized message to any kind of state change, using the proposed method the amount of dynamic information is independent of the complexity of the given optical device.
Of course, this kind of information reduction is not gratuitous: its price is paid at the PCE, since the path computing algorithm has to take more computation steps. The PCE relies on the following associative arrays:

- head[edge] is the head node of the edge,
- tail[edge] is the tail node of the edge,
- in_edge[node] is the set of ingress edges,
- out_edge[node] is the set of egress edges,
- total_cap[edge] initial capacity of the edges,
- sum_cap[edge] actual allocated capacity of the edges.

And the actual amount of free capacity that can be allocated on a link is defined by the free_cap(e) function:

```plaintext
function free_cap(e)
  if (sum_cap[e] > 0) // already switched
    return total_cap[e] - sum_cap[e]
  out_n=0
  in_n=0
  for_each (i in out_edge[head[e]])
    if (i != e and sum_cap[i] > 0)
      out_n++
  for_each (i in in_edge[tail[e]])
    if (i != e and sum_cap[i] > 0)
      in_n++
  if (out_max[head[e]] <= out_n or
      in_max[tail[e]] <= in_n)
    return 0
  return total_cap[e]
```

This way, the evaluation of the free_cap(e) function increases the complexity of the path computation, at the same time, the task of the optical devices will be simplified as they do not need to identify the additionally affected WG edges. In fact, we have moved the functionality of the identification of affected WG edges from the optical devices to the PCE, resulting in a simplified communication between these two network elements.
4.1.2 Generalized Protection Formula (GPF)

Although by applying the presented RBTA method we can significantly reduce the amount of LSA messages in the network, the RBTA has the serious shortcoming that internally it does not support protection, i.e., conditional (failure-dependent) switching rules cannot be expressed by it. Thus we have extended the original RBTA to support protection. However, in this extended RBTA version we do not want to constrain the set of applicable protection schemes. Hence we need to define a unified and general way of describing any conventional protection scheme.

**THESIS 1.2 ([C2]):** I have introduced the GPF (Generalized Protection Formula) function, that provides a unified way to describe the switching and allocation requirement of paths realizing the network connections. This generalized description provides a simple and integrated way for network operators to deploy any of the Conventionally used protection schemes in the network and use them simultaneously.

The GPF function has taken ideas from Partial Path Protection [11] and Subgraph Routing [12]: by means of GPF we can describe in a unified manner both dedicated and shared, both link and segment protections. The key of this flexibility is the fact that GPF assigns the switching states of nodes for each connection failure-dependently. GPF has a 3-dimensional domain:

- $C$ is the set of connections,
- $V$ is the set of WG nodes, and
- $F^*$ is the set of considered failure scenarios (including no failure state).

The function assigns for each item of the domain a value pair from the set of ingress and egress edges extended by the indefinite item $e^n$ ($E^* = E \cup \{e^n\}$):

$$GPF : C \times V \times F^* \rightarrow (E^* \times E^*)$$

The meaning of $GPF(c, n, f) \rightarrow (e_1, e_2)$ function is that the connection $c$ is switched in node $n$ in case of failure state $f$ from edge $e_1$ to edge $e_2$. In other words, the GPF mapping defines the switching action of each network device for each connection in case of any failure.

By means of GPF we can describe the network states even if we use simultaneously protection schemes that have different scope and shareability. Schemes that handle multiple simultaneous failures or apply partial protection to the connections can be traced back on a GPF applied to a transformed domain of the $C, V, F^*$ triplet.
Regarding the scope of the protection, the GPF can handle end-to-end, segment and shared protections. In case of end-to-end protection we group the failure cases of $F^*$ into two sets, the one set contains failure cases that require the backup path to be used, the other set contains the remaining failure cases. In case of segment and link protection, there will be created multiple groups, each containing failure cases triggering a given backup path.

In the dimension of shareability we distinguish three alternatives: 1+1, 1:1 and shared protection. In 1+1 protection all the nodes are configured statically for every $f \in F^*$, in case of 1:1 protection, the realization is the same, except for the source and destination nodes of the protection or protection part. These exceptional nodes switch to the backup path only in case of failures described by previously defined groups. The shared protection differs from the latter one in the manner, that not only the source and destination nodes, but also all the intermediate nodes are failure case sensitive, meaning that as long as no failure to react to occurs, these intermediate nodes remain not configured.

Further dimension might be the amount of protection, however, this dimension can be eliminated: if the protection path can carry only $x\%$ ($x < 100$) of the traffic, two traffic flows are routed instead of one: the one carries $x\%$ of the original traffic with protection, the other one carries the rest ($100 - x\%$) of the traffic without any protection.

The GPF describes one-failure-survivable schemes. However, the failure survivability, expressing how many network devices can simultaneously fail without breaking the connection, is also a dimension that can be transformed into single-failure-survivability and other constraints. By extending the concept of Shared Risk Link Groups [8], even originally independent devices are allowed to form a risk group and dual link or node failures can be transformed into single SRLG failures which already can be expressed in GPF.

### 4.1.3 Rule-Based Topology Advertisement supporting Protection (RBTA-P)

In Thesis 1.1 we have proposed a method (RBTA) to advertise the network state efficiently after resource allocations and deallocations. In Thesis 1.2 we presented a formula (GPF) that can describe different protection schemes in a unified manner. Our aim is to extend the RBTA method by means of GPF to support link state advertisement even if the connections are protected.

**Thesis 1.3 ([C2]):** Based on Theses 1.1 and 1.2 I have proposed the method RBTA-P (RBTA supporting Protection) for state advertisement and information processing, which provides a compact and lossless description of the switching state of the network in any failure-dependent – including failure-free – cases. The signaling and data overhead of the method is proportional to the number of state changes and the average number
of considered failure scenarios along the working paths.

In case of RBTA-P we have to answer the same questions as in case of RBTA:

- What information should the controller of the crossconnect advertise?
- How should the TED store the data?
- How can the PCE process the data in case of routing?

The invariant parts of the advertised data, i.e., the switching rules of the ports, are the same as in RBTA. The dynamic advertised information of RBTA has to be extended, as we have to define what failure scenarios activate the given switching. Assuming that the set of activating failure states is described by SRGs, the basic LSA message is:

\[ LSA(i) : (\text{link	extunderscore}id, \Delta \text{cap}, \text{SRGs}) \]

In the SRGs attribute the failure states are coded by simple enumeration by enumeration of the inverted set of activating failure states. The usage of inverted set is favorable in cases when the switching applies to almost every failure cases (e.g., switching of the default path). By using this inverted set we manage to constrain the amount of the advertised information to be proportional to the average number of nodes along the working paths.

In the TED we have to store the amount of allocated capacity per failure state, hence, instead of the 1-dimensional \( \text{sum	extunderscore}cap[e] \) array of RBTA, in RBTA-P uses the 2-dimensional \( \text{sum	extunderscore}cap[e][f] \) array, where the variable \( f \) takes its values from set \( F^* \).

In RBTA-P, compared to RBTA, the \( \text{free	extunderscore}cap(e,f) \) function of the PCE is similar to the \( \text{free	extunderscore}cap(e) \) function of the RBTA. The only difference in \( \text{free	extunderscore}cap(e,f) \) is that it refers to the \( \text{sum	extunderscore}cap[e][f] \) array instead of the \( \text{sum	extunderscore}cap[e] \) array.

In case of protection we deal with a set of failure scenarios \( (F_x \subseteq F^*) \) instead of individual failure scenarios. We have to define the state of the network links for the set \( F_x \subseteq F^* \) instead of a single \( f \in F^* \). This link state can be evaluated as the intersection of edge states corresponding to any \( f \in F_x \) failure scenario. In case of the free capacity metric, this intersection is given by the minimum of the \( \text{free	extunderscore}cap(e,f) \) values:

\[ \text{Free	extunderscore}Cap(e,F_x) = \min_{f \in F_x} (\text{free	extunderscore}cap(e,f)) \]

There are more benefits of the RBTA-P solution. First, the PCE is aware of “all” the resource allocation and switching information necessary to make routing decision. In case of failure state change the re-advertisement of the modified switching state of the network is not needed, as it is already known by the PCE. Second, by means of this description we can
simultaneously use each of the conventionally widespread protection schemes without any
protocol extension – dedicated or shared; end-to-end, segment or link protection – since the
shareability information is advertised in a unified manner. Last but not least, the advertised
state information can be used as control information in case of failures, thus the load of the
control channel can be reduced significantly.

4.2 Multidomain p-cycle protection

In the previous section we gave an overview on how to advertise efficiently intra-domain
topology and link state information. By means of these solutions we can achieve that
the PCE has always accurate information on the network state. Based on this accurate
information the PCE can efficiently compute paths for the connections, even with protection.

In case of inter-domain routing, however, we face different problems. For example, the
operators of the domains do not want to share internal topology with competing operators.
Furthermore, there are failure risks that are considered to be independent, however, in fact
they are correlated (SRLG overlaps). Due to these problems the provision and the protection
of inter-domain connections is a far more complex task than of intra-domain connections.
These problems also imply that many conventional protection schemes become unfeasible in
multidomain networks.

As a result of our research, we have proposed methods for inter-domain topology adver-
tisement. By means of these methods we can provision inter-domain connections that take
into account predefined QoS parameters. However, the proposed methods do not consider
resilience. To overcome this shortcoming, in Thesis Group 2 we propose a p-cycle-based
protection scheme for QoS guaranteed inter-domain connections.

There are more benefits of using p-cycles for protection. p-Cycles are resource efficient
and provide fast protection switching. Moreover, the protection paths in the p-cycle scheme
are preconfigured, and in case of inter-domain protection this is a very important property.
Due to this preconfiguration the operators of the different domains can assign the protection
paths in advance. Thus, SRLG overlaps can be avoided due to such offline management
negotiations.

In this Thesis Group (Sect. 4.2) I focus on two questions. First, we describe how to
realize p-cycle protection in inter-domain environment. Second, we examine how to balance
between resource consumption and connection availability.
4.2.1 Constructing multidomain $p$-cycles

In case of inter-domain connections, the intra-domain parts of the connections can be protected easily. Our aim is to provide $p$-cycle-based protection for the traffic on the inter-domain network links. This way we can protect inter-domain connections along their entire length.

**Thesis 2.1** ([J1, J2, C6, C7]): *I have proposed the Multidomain $p$-Cycle (MDPC) solution protect traffic routed on inter-domain network links. MDPC, unlike many other protection schemes, is feasible in inter-domain environment."

![Diagram of network topology](image)

Figure 1: Supposed two level topology information for constructing multidomain $p$-cycles

We assume that a multidomain network is modeled by a two-level topology as shown in Fig. 1. At the higher level only the border nodes and inter-domain links are known definitely. The internal topology of the domains is hidden at the higher level, only *virtual* links between the border nodes denote the connectivity. At the lower level, each domain is aware of its own internal topology, however none of them has any information about the network outside the domain.

In order to apply MDPC protection in the network we have to define the candidate inter-domain cycle set. In conventional, intra-domain context, there are multiple cycle search algorithms. If we want to apply these algorithms in inter-domain context, we have to model our network as a graph where the domains are represented by the nodes of the graph whereas the inter-domain links are the edges of the graph. This graph can be constructed at the higher hierarchy level easily.

Using the conventional cycle search algorithms we can enumerate the candidate $p$-cycles at the higher hierarchy level. However, in these enumerated cycles the domains are modeled
by nodes. Hence, within the domains, the on-cycle and straddling inter-domain links of the 
p-cycle have to be connected. This intra-domain connection is the task of the domains.

For each inter-domain p-cycle within each domain we define two types of border nodes:

**CBN** is a border node where an on-cycle link ends;

**SBN** is a border node where a straddling link ends.

To resolve the intra-domain connection of the inter-domain p-cycle, in each SBN we have to originate two paths leading to the two CBNs. Furthermore, the two CBNs also need to be connected. The links along these paths define the $E_1$ set. Next, we have to define how much capacity should be allocated on these links. We examine the two outgoing paths in each SBN: those links that are common in these two paths are put into set $E_2$. Afterwards, on the links of $E_2$ we must allocate 2 units of capacity, whereas on the remaining members of $E_1$ it is sufficient to allocate only 1 unit of capacity. Finally, in order to support higher level p-cycle protection assignment, the cost of this intra-domain p-cycle resolution has to be advertised to the higher hierarchy level.

The selected paths between the SBNs and CBNs influence the resource requirement and the provided availability of the p-cycles. The simplest solution to define these paths is to route them on the shortest (or most reliable) paths. This way we get the Most Reliable (MR) intra-domain resolution of MDPCs. However, we have elaborated other resolutions as well, and two of them are presented in Sect. 4.2.2.

### 4.2.2 Intra-domain resolution of inter-domain p-cycles

In Sect. 4.2.1 we have presented a method that produced inter-domain p-cycles as a result of cooperation of the lower and the higher hierarchy level of the network. We have seen that in this procedure the role of the domains at the lower level is important, since they have to resolve the connections between the inter-domain on-cycle and straddling links. The simplest way of intra-domain resolution of MDPCs (MR) does not focus on the sparing resource usage. Moreover, there are methods that result in more available p-cycles than the MR method does. The MR method is not optimal either from resource usage or from availability point of view.

**THESIS 2.2 ([J1, C6]):** I have proposed the Least Cost (LC) and Ring Based (RB) methods for intra-domain part resolution of inter-domain p-cycles. These methods require less resources (LC) and provide higher availability (RB) than the original Most Reliable (MR) solution.
Figure 2 shows possible resolutions of the intra-domain part of MDPCs.

When we minimize the resource consumption, we can use the LC method. The steps of the LC algorithm are as follows:

1. Initially, we register each border node into the set $S$ as subgraphs containing only a single node.

2. For each subgraph $s_i \in S$ we search for its nearest $s_j \in S$ neighbor and compute their distance $d(s_i, s_j)$.

3. We select the closest $(s_i^*, s_j^*)$ pair with minimal $d(s_i^*, s_j^*)$. Let $R^*$ denote the shortest path between $s_i^*$ and $s_j^*$. Delete $s_i^*$ and $s_j^*$ from the set $S$ and insert $s_x = s_i^* \cup s_j^* \cup R^*$ into $S$ instead of them.

4. If there is still more than one member in $S$, go to step 2.

5. As a result, $S$ will contain only one subgraph $s_{res}$ which we were looking for.

The complexity of the LC method is polynomial. Having $|V|$ nodes, $|V_B|$ border nodes and $|E|$ edges in the domain, in the first iteration we perform $|V_B|$ times Dijkstra's algorithm, totally in $O(|V_B| \cdot |E| \cdot \log(|V|))$ steps. The algorithm has $|V_B| - 1$ iterations resulting in an overall complexity of $O(|V_B|^2 \cdot |E| \cdot \log(|V|))$. 
We also may aim to improve the connection availability. This improvement may be required as the inter-domain $p$-cycles are usually much longer and less reliable compared to the intra-domain $p$-cycles. The Ring Based (RB) solution [11] tries to chain up the CBNs and the SBNs on a ring. This way there are two alternate paths (default and backup) provided between each border node pair.

The calculation complexity of the method is negligible if the $p$-cycles inside the domain are already enumerated. In this case we have to choose a proper candidate $p$-cycle that contains all the CBNs and SBNs. If such a cycle cannot be found, we choose a candidate that contains most of the CBNs and SBNs, and we connect the remaining CBNs and SBNs to this cycle by a single path.

4.2.3 Comparing different MDPC solutions

The different intra-domain resolutions (LC, MR, RB) are applied for MDPC based protection schemes named CIDA and CIDED. CIDA protects both inter-domain links and intra-domain links with $p$-cycle, whereas CIDED uses dedicated protection to protect intra-domain links. As a reference, we have evaluated the end-to-end dedicated protection (Ded. E2E) (although in multidomain networks it is not feasible) and the case with no protection (No prot.).

![Relative Resource Requirement of different protection schemes](image_url)

Figure 3: Relative resource requirement of different protection schemes

Figure 3 shows for different networks the relative resource requirement of protection schemes compared to the ‘no protection’ case. It can be seen that LC uses less resources than MR.

Figure 4 shows the ratio of connections having higher availability than a predefined value (on x axis). It can be seen that the RB method, applied both in CIDA and in CIDED, significantly increases the connection availability.
4.3 Efficient cycle search and availability evaluation methods for $p$-cycles

There are several favorable properties of the $p$-cycle protection scheme. These properties motivated us to use $p$-cycles to protect inter-domain connections. During the research, however, we have faced problems, that were not only multidomain protection specific, but also general problems related to $p$-cycles. In Thesis Group 3 we examine these problems.

I have proposed procedures to enhance the efficiency of $p$-cycle employing systems. This thesis group contains three theses. Thesis 3.1 presents an efficient way of enumerating candidate $p$-cycles. In Thesis 3.2 an efficient and fast method of availability evaluation is proposed for $p$-cycle protected connections. Finally, Thesis 3.3 analyzes the accuracy of the serial-parallel availability estimation heuristic applied for $p$-cycle protected connections.

In sections 4.3.2 and 4.3.3 regarding the $p$-cycles we assume that:

- There are only link failures (node failures are transformed into link failures).

- We protect only against single failures.

- The protection switching transient is negligible short compared to the mean time to repair. The availability of a connection can be directly derived from the availability of the network links.

- The protection switching mechanism is based only on local (on-cycle and straddling) link states.
• Changing the state of any link from Down to Up cannot imply the state change of any connection from Up to Down.

• The protection assignment is predefined, failure priority based. In case of multiple failures this assignment does not depend on the sequence of failure occurrence / detection.

4.3.1 The Exhaustive Grow p-cycle search algorithm

There are many known p-cycle searching algorithms. The Grow and Expand algorithms [13] are efficient heuristics. However, they do not consider that replacing a link to a disjoint route may restrict the replacement of another link in the p-cycle, therefore the outputs of the algorithms depend on the sequence of the on-cycle links. This way these algorithms may omit to enumerate some efficient candidate p-cycles.

Thesis 3.1 ([C6, JH3]): I have proposed the Exhaustive Grow algorithm, as an extension of the Grow algorithm, to detect a larger (at least equal) set of candidate cycles that is still smaller than the theoretically maximal set. Compared to Grow, using the cycle set of Exhaustive Grow we can protect the connections by less resources, and compared to the theoretically maximal set, Exhaustive Grow produces less cycles, thus the protection assignment requires less computational steps.

Unlike the Grow algorithm, the Exhaustive Grow considers all the possible sequences of the on-cycle links which results in a larger set of p-cycles.

The algorithm can be represented in pseudo-code as follows:

ExhaustiveGrow(CycleSet pcs, Graph G) {
    while (Cycle p = pcs.next()) {
        Graph I = Graph();
        for_each(Node n in p) I.add(n);
        for_each(Link l in p) I.add(l);
        for_each(Link l in p) {
            Route r = ShortestPath(l.head, l.tail, (G \ I) ∪ {l.head, l.tail});
            if (r != null) {
                Cycle p’ = p;
                p’.remove(l);
                p’.add(r);
                pcs.add(p’);
            }
        }
    }
}
To confirm that the Exhaustive Grow finds a larger set of cycles than the Grow and Expand algorithms did, we have taken existing network topologies and also artificially created ones and have searched for cycles in them. Simulation results (Fig. 5) have shown that the Exhaustive Grow algorithm produces more candidates than the two reference algorithms. Using the produced cycle sets in the CIDA algorithm we have assigned $p$-cycle protection to the working capacity in different networks. The results show that compared to the Grow algorithm, with the Exhaustive Grow we can achieve a resource usage reduction of more than 5% on average in the selected networks (Fig. 6).

We have evaluated the performance of Exhaustive Grow and set it against the optimal, ILP-based solution. The results of the ILP-based solution are taken from [13]. In Table 1 ‘Redundancy’ denotes the resource requirement of the protection relative to the resource requirement of the working paths of the connections. ‘% Diff’ denotes the relative additional resource requirement compared to the optimal solution.

The results in Table 1 show that although the Exhaustive Grow algorithm produces only
approx. twice as many cycles as the Grow, using its cycle set of Exhaustive Grow the CIDA can approximate the optimal solution significantly.

4.3.2 Availability evaluation of p-cycle protected connections

If we evaluate the availability of p-cycle protected connections using the Serial-Parallel heuristic, we have to count with approximation error. To get the exact availability of the connection, we have to rely on the key element method [9].

In connection \( conn \) of length \( n \) the links are protected separately. The values \( a_i \) and \( b_{i,j} \) \((1 \leq i \leq n)\) denote the binary state (Up or Down) of the \( i \)th working link and of the \( j \)th protection link in the protection of the \( i \)th working link, respectively. If we split the connection into two parts after the \( i \)th link, we get \( conn_i^H \) and \( conn_i^T \) as the head and the tail of the connection. Value \( X_i^H \) denotes the state of \( conn_i^H \). This way, the connection availability is given by the probability \( P(X_n^H) \). We also define the farthest overlap (\( L_{\text{max}} \)): it is the maximal distance \( k - i \) in the connection so that the set of resources (set of the working and backup links) of the \( i \)th and \( k \)th working link have any link in common (\( conn_i^H \) and \( conn_k^T \) have common link).

**Thesis 3.2** ([CS]): I have proposed a recursive probability calculation method for getting the exact availability of a connection protected by p-cycles. The calculation complexity of the method is of \( O(n \cdot 2^{L_{\text{max}}+1}) \).

The availability of the connection \( P(X_n^H) \), using the key element method, is defined
recursively:

\[ P(X_i^H) = P(a_i) \cdot P(X_{i-1}^H | a_i) + P(\overline{a_i}) \cdot P(X_{i-1}^H | b_{i,1}, b_{i,2}, \ldots, b_{i,K_i}, \overline{a_i}) \cdot \prod_{j=1}^{K_i} P(b_{i,j} | a_i), \quad (4.1) \]

where the terminating value for any \( G \) condition is

\[ P(X_0^H | G) = 1 \quad (4.2) \]

This recursive calculation method requires theoretically \( 2^{n+1} - 1 \) steps, however, practically we can find a much lower upper bound.

We can trace back \( P(X_n^H) \) to weighted sum of conditional availabilities of \( X_i^H \). In the formula of \( P(X_n^H) \) there will be \( P(X_i^H | G) \) terms of different \( G \) conditions. However, each of these \( G \) conditions may refer to a reduced set of links: the referred links must be members both of the connection head (from the first to the \( i \)th link) and of the connection tail (from the \((i+1)\)th to the last link). This set of links is identical to the set of overlapping links. If the farthest overlap is not greater than \( L_{\text{max}} \), then, at recursion level \( i \), the different \( P(X_i^H | G) \) conditional availabilities cannot contain more than \( 2^{L_{\text{max}}+1} \) different \( G \) conditions.

As the depth of the recursion is \( n \) and its maximal width is \( 2^{\text{max}+1} \), altogether, the calculation complexity of the method is \( O(n \cdot 2^{L_{\text{max}}+1}) \).

### 4.3.3 Approximation error of the Serial-Parallel estimation

Although the serial-parallel (S-P) calculation method offers a very fast way for retrieving connection availability metric from basic link availability metrics, (having \( n \) links its complexity is \( O(n) \)) it does not take into account the link overlaps, i.e., those elements that are common in different series. This shortcoming of the algorithm may lead to approximation error and inaccurate estimations.

**THESIS 3.3 ([C9]):** For cyclic protected connections I have defined the approximation error of the S-P method, the bounds of this error and with certain restrictions its limit value as the link unavailabilities tend to 0 in case of overlapping working or backup links.

In the results we use the notation of Sect. 4.3.2 and the following annotation:

- \( U_{\text{ACC}} \): is the exact unavailability of the connection.
- \( U_{\text{SP}} \): is the unavailability of the connection estimated by the S-P method.
- \( U_{\text{SP}}^H \): is the S-P unavailability of the connection head. \( U_{\text{SP}}^T_i \) is defined similarly for the connection tail.
$U_{SP|G}$: conditional $S$-$P$ unavailability with $G$ condition. Similarly we define $U_{SP^H_{i}|G}$ and $U_{SP^T_{i}|G}$ for the conditional connection part unavailabilities.

$DIV_U$: relative accuracy of unavailability approximation ($DIV_U = \frac{U_{acc}}{U_{SP}}$).

We assume that condition $G$ ($G_1$ and $G_2$ similarly) defines conditions only for the state of the overlapping links. This way we can define the maximal difference between conditional unavailabilities of the connection parts (head or tail):

$$
\delta^H_i = \max_{G_1,G_2} \left| U_{SP^H_{i}|G_1} - U_{SP^H_{i}|G_2} \right|
$$

$$
\delta^T_i = \max_{G_1,G_2} \left| U_{SP^T_{i}|G_1} - U_{SP^T_{i}|G_2} \right|
$$

$$
\delta_i = \max(\delta^H_i, \delta^T_i)
$$

Using this notation we can state:

$$DIV_U \leq 1. \quad (4.3)$$

This is denoting that the $S$-$P$ approximation never overestimates the availability. Furthermore, if the head and the tail connection parts do not contain overlapping links we can define a lower bound of $DIV_U$ (Eq. (4.4)) and the limit value of $DIV_U$ as the the link unavailabilities tends to 0 (Eq. (4.5) and Eq. (4.6)):

$$DIV_U \geq 1 - \frac{1}{2} \delta_i. \quad (4.4)$$

By increasing the value of the link availability metric in the network, i.e., $\forall e: P(S(e) = 0) \leq \varepsilon$, we have shown that the deviation of the calculated and the accurate connection unavailability converges to a defined value. If there are no mutual working link overlaps in the connection:

$$\lim_{\varepsilon \to 0} DIV_U = 1 \quad (4.5)$$

If the protections of neighboring working links mutually overlap each other, and we denote the working state of these working links by $S_1$ and $S_2$, their failure state by $\overline{S_1}$ and $\overline{S_2}$, whereas their unavailability by $U_1$ and $U_2$:

$$\lim_{\varepsilon \to 0} DIV_U = 1 - \frac{1}{2 + \lim_{\varepsilon \to 0} \frac{U_{SP^T_{i}|S_1,\overline{S_2}}}{U_1} + \lim_{\varepsilon \to 0} \frac{U_{SP^H_{i}|\overline{S_1},S_2}}{U_2}}. \quad (4.6)$$

21
5 Practical Application

The presented results can be applied to multiple fields of telecommunications. First, the results of Theses 1.1 and 1.3 can be applied to the optical layer of modern telecommunication networks. Using current technologies, the reconfiguration of the optical layer is problematic, and one of the reasons is the huge amount of state information to be advertised after each reconfiguration. The proposed schemes make this advertisement easy to realize, thus a more flexible way of optical layer reconfiguration becomes feasible.

The GPF description of protections (Thesis 1.2) may be applied in heterogeneous systems where multiple protection schemes are used simultaneously.

In multidomain networks, the results of inter-domain p-cycles are of higher importance (Theses 2.1 and 2.2). Multidomain p-Cycles provide a simple and feasible way of protecting inter-domain traffic. In [J1] there are solutions proposed how to realize MDPC with current technology (e.g., BGP-4).

Regarding the availability evaluation, for example, we can construct an algorithm that finds the most available, p-cycle protected path between two network nodes, based on the A* algorithm applying the results of Thesis 3.3 as upper and lower bound estimation.

6 Acknowledgment

I would like to thank to my supervisor, Tibor Cinkler, for all his kind help. Many thanks to my mother and father for their permanent motivation. Finally I would also like to thank to the staff and employees of the HSN Laboratory and Department of Telecommunication and Media Informatics for their support even after I finished the postgradual studies and I have left the department.
References


Publications

Journal Papers


Journal Papers in Hungarian


Conference Papers


Other Publications


27
