20

Redundancy Domains - a novel Approach for Survivable Communication Networks

Andreas Iselt
Lehrstuhl für Kommunikationsnetze, Technische Universität München
80290 München, Germany, tel. +49 89 289-23504, fax. -23523
iselt@lkn.e-technik.tu-muenchen.de

Abstract

To maintain a high availability of communication networks, traditionally two separate mechanisms are used, redundancy within network elements and network-wide redundancy. Usually these mechanisms are applied independently. In this paper a new technique is presented, which allows the redundant provision of simple non-redundant network elements instead of internally redundant elements. For that purpose new devices are proposed, which implement the protection functionality outside the usual network elements. Two different types of these devices are described, that allow non-stop, hitless operation in the case of a failure, using 1+1-redundancy or diversity techniques.

Keywords

Reliability, Survivability, Redundancy, Diversity

Broadband Communications P. Kühn & R. Ulrich (Eds.)

1 INTRODUCTION

The technological progress in the fields of semiconductors, communications and networking allows modern broadband networks to transmit growing amounts of data. The high bandwidths of links and crossconnect systems lead to a reduced number of both of them with constant or enhanced network capacity. Unfortunately, faults in these systems affect an increased number of end-to-end network connections and lead to more loss of data. Consequently, it is important to maintain a very high availability for every part of the communication network.

Conventional approaches achieve this goal using multiple mechanisms. One mechanism is to keep the availability of the network elements at a high level. This is accomplished using redundancy mechanisms inside the network elements (e.g. redundant switch matrices). Another mechanism to increase the network survivability, is to use network wide procedures (e.g. automatic protection switching). Usually these different methods work uncoordinated.

In this paper a new approach is shown, that uses simple network elements and protects them externally to achieve a high reliability. The partitioning of the network in so-called redundancy domains with individual fault tolerance competence is the basic concept and has already been claimed for patent [St95].

Only few of the currently available techniques have the property of hitless operation in the case of failures. However, these methods require 100% redundancy of the protected entities. To achieve a reduction of the required bandwidth, diversity techniques are proposed in this paper, that are mainly known from higher protocol layers.

2 CONVENTIONAL APPROACHES

To minimize the effect of failures in communication networks and to be able to guarantee a continuos operational state, different fault tolerance mechanisms have been proposed and partly implemented. Breakdowns of network elements (e.g. crossconnects, switches, multiplexers) can be reduced by providing internal redundancy in the network element. In the case of link failures or network element failures, network-wide procedures can recover from the error.

2.1 Network Element Reliability

Network elements are composed of a number of different modules. Figure 1 shows the structure of a typical network node, composed of line interface modules, switch matrices, controller and power supply. Each module may fail independently of the others, although the node may also fail totally, e.g. as a consequence of fire or sabotage. To be able to tolerate the failure of single entities in the node, spare modules are provided.

In nodes with high reliability demands, the switch matrix is usually duplicated, and running in hot standby mode with both matrices working concurrently, thus allowing a hitless recovery from failures (e.g. [FFGL91]). Other approaches use the effect, that the switch matrix is often spread over several switching modules and

form switching networks within the network element, that may be protected using m:N-redundancy for the switching modules ([It91], [YaSi91]). These fault tolerant switch matrices cannot operate hitless, since a reconfiguration is necessary after the detection of faults. Power supplies are often provided redundantly even in simple low cost switching systems, since they do not require very much effort and offer a simple means to improve a nodes availability. For highly reliable systems, redundant controllers may be provided as sketched in Figure 1. If the hardware of the controller is foreseen redundantly, a synchronization of micro instructions is necessary to achieve non stop operation in the case of a breakdown of one controller. Software errors cannot be overcome even with multiple redundant hardware modules. To tolerate these, separately developed software would be necessary. Micro instruction synchronization isn't possible with this kind of redundancy, thus leading to interruptions when one processor takes over the operation of the other. Very important for the availability of network elements are the interface modules. These modules cannot be protected by automatic mechanisms within the network element and are protected together with the network links, as described later.

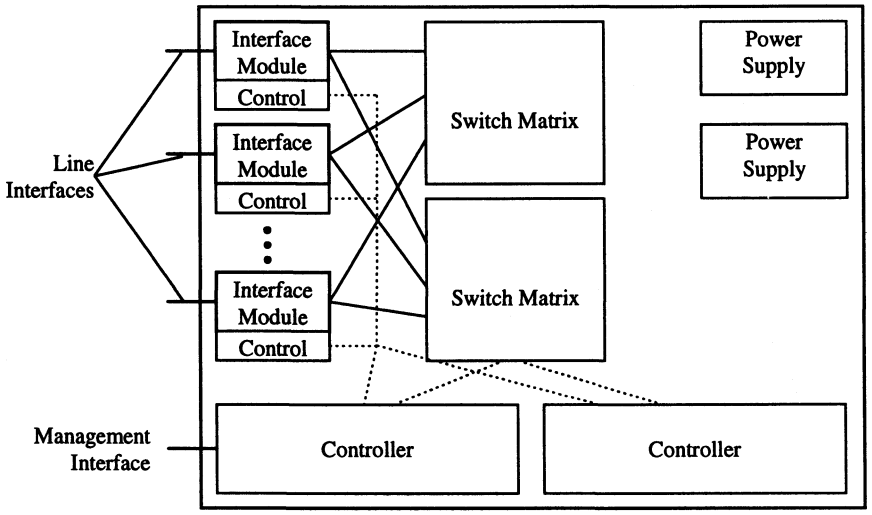


Figure 1 Internally redundant Crossconnect System.

Hence, it is clear, that the design of reliable network elements is not trivial and requires an fairly increased effort compared to simple, non-redundant network elements, like they are mainly used for private network applications.

2.2 Network Survivability

In addition to the methods which increase the network elements' availability internally, a large variety of network wide approaches to maintain a high level of survivability exists. They are used to recover from link failures or from node failures, if the internal mechanisms are not able to fix the problem. Figure 2 shows a common classification of recovery techniques, based on [I.311] and [Ed95].

The advantage of the "Centralized Rerouting" approaches is, that the central network management system has a global view over the network and may achieve a new optimal routing in the case of failures. Due to the high signaling effort and the difficult calculation of the new routing these approaches are rather slow and may lead to enormous loss of data or even drop of calls.

The "Selfhealing Networks" implement distributed algorithms to reroute the connections in case of failures. This distributed determination of new paths speeds up the recovery, but still a considerable loss of data and call dropping occur.

	Protection *Redundant paths predefined*	***Restoration*** *Protection paths determined during recovery*
Distributed	Protection Switching, Diversity	Selfhealing Networks (distributed rerouting)
Centralized	-/-	Centralized Rerouting

Figure 2 Classification of Fault Recovery Techniques

The fastest recovery is possible with the "Protection Switching" methods. Using the terms "1:1", "1:N" and "m:N", methods are classified, that have 1 or N protection paths for 1 or m working paths. If a working path fails, it may be switched to a protection path. Therefore not more than N working paths may be recovered at the same time. In the case of no failure having occurred, the protection paths may be used for low priority traffic. Since these methods require the source and the sink of the protected path to be switched to the protection path, requiring an amount of time (usually some milliseconds), this procedure cannot operate hitless and a break in the transmitted signal still occurs.

With 1+1-redundancy, data is always transmitted on two redundant paths. At the receiving side one path is chosen depending on the failure state of the two paths. This approach has the advantage, that only the receiver has to switch to the protection path and no switching operation is necessary at the sender. But therefore 100% redundancy is necessary and no low priority traffic may be transmitted on the protection path.

With diversity techniques it is possible to reduce the required bandwidth by distributing the traffic on multiple links with lower bandwidth and protecting only against single link failures with support for hitless recovery. A more detailed investigation is given later in this paper. An overview on diversity techniques can be found in [GuKa97].

2.3 Evaluation of known Survivability Techniques

In the redundancy approaches presented up to now, the following disadvantages may be identified. They do not offer an integrated approach for network redundancy and network elements' redundancy. The application of the fault recovery techniques on an

freely chosen partition of the network is not supported. Most approaches do not provide non-stop operation in case of failures. The integration of redundancy in the network elements may be uneconomical. Further, no existing approach integrates fault recovery techniques with fault correction techniques, although both may be based on a common redundant path.

3 NOVEL REDUNDANCY TECHNIQUE

3.1 Redundancy Domains

To circumvent the prerequisite, that the protection switching operation has to be implemented at topological and physical predefined places (in general in the network elements) the concept of partitioning the network in redundancy domains is proposed. This makes it possible, to provide the whole network element redundantly. A redundancy domain is a subnetwork consisting of two or more redundant network elements, links or subnetworks. These redundant parts coexist independently of each other. They are connected only at so called redundancy domain gateways (RDGs) at their common border. In these new devices the fault detection and protection switching functions are implemented.

Figure 3 shows an example for the redundancy domain principle. Domain D2 is composed of two network elements (e.g. crossconnect systems). At the border of the domain, the redundancy domain gateways connect the two redundant elements. Domain D1 shows, how the same approach can also be applied to network links.

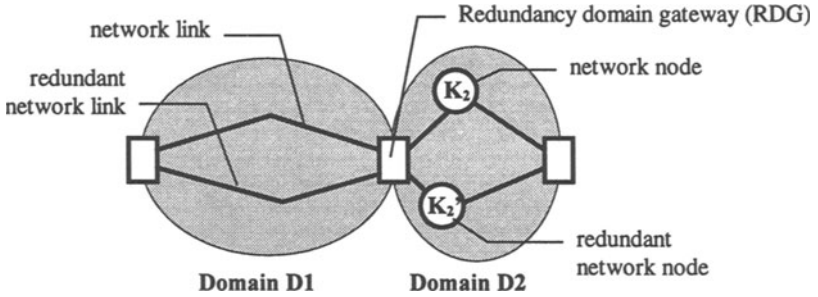
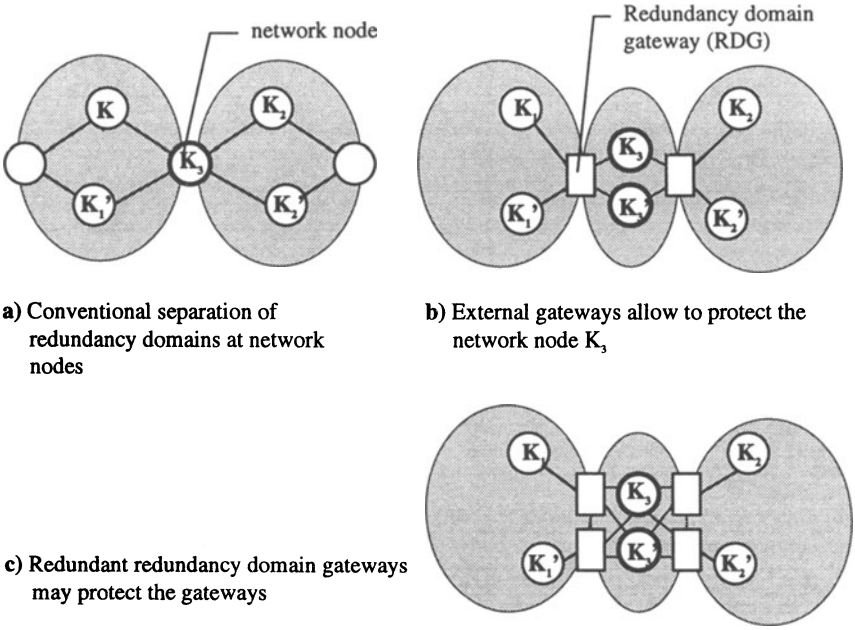


Figure 3 Redundancy domain principle

The use of redundancy domains as protection mechanism is equal to subnetwork protection in the sense of [G.805]. The main difference is, that up to now the protection switching function has always been collocated with the normal switching function. In general the switch matrix of existing network elements is also used for the protection switching operation. With the new approach the protected domains may cover any part of a network and the protection function has not to be physically located at existing network elements (e.g. crossconnect systems). Thus, the network element itself can be protected without the need for internal redundancy. Low-cost, non-redundant devices (e.g. equipment for private-network application) may be util-

ized. Furthermore, network links can also be protected at other points than crossconnect or switching systems, allowing even to split long links in several domains.

In principle the redundancy domain view can also be applied to existing redundancy approaches. In Figure 4a an example is shown, where a network connection is protected using redundant paths with separate network elements (K_1' protects K_1, K_2' protects K_2). In this case, the protection function has, up to now, always been collocated with a network node (K_3). The protection of this network node poses a problem, since redundancy may only be integrated internally and no protection against total network element failure is provided.



a) Conventional separation of redundancy domains at network nodes

b) External gateways allow to protect the network node K_3

c) Redundant redundancy domain gateways may protect the gateways

Figure 4 Node protection with redundancy domains

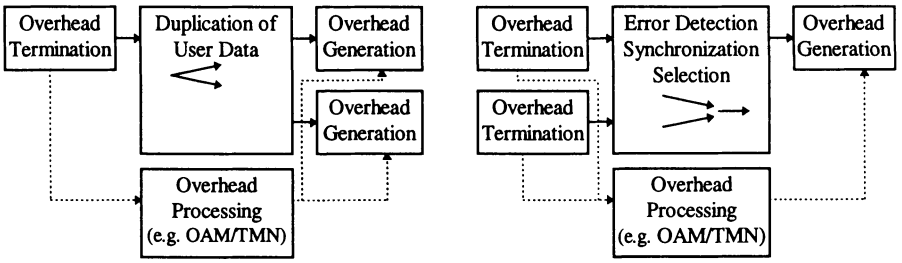
With the use of the new RDG devices, it is possible to build an own domain around the network element, which can now be provided redundantly (Figure 4b). Obviously the problem is now transferred to the gateway elements which now represent so called single points of failure. With the use of redundant RDGs this problem can be solved (Figure 4c).

Within the domains, in principle any type of protection may be applied. In this paper only protection mechanisms with hitless operation are investigated, namely 1+1-redundancy and diversity. For both types of protection the corresponding new RDG devices are described in the following sections.

3.2 Redundancy Domains with 1+1-Redundancy

With 1+1-redundancy, the data streams are duplicated when entering a domain and transmitted on two different paths. Before leaving the domain one data stream has to

be selected. The redundancy domain gateway for 1+1-redundancy is called DS-gateway (Duplicate/Select). Block diagrams for the duplicator and the selector part of the gateway are shown in Figure 5. It is important to notice, that besides the duplication, synchronization and selection of the user data, the protocol overhead (e.g. OAM messages) also has to be processed.



a) Block diagram of duplicator device **b)** Block diagram of selector device

Figure 5 Block diagrams for Duplicator and Selector

The duplication function is rather simple. Protocol overhead from incoming data streams is stripped off and the user data is forwarded to two outputs where new protocol overhead is added. Depending on the protocol layer in which the duplication is realized, more or less effort for the overhead processing has to be spent. Duplication in the physical layer for example may be implemented using passive optical splitters, whereas duplication in the ATM layer requires the processing of all the layers beneath. On the other hand, using higher layers makes it possible to selectively duplicate only logical connections with high reliability constraints.

For the selection operation also the overhead processing functions have to be provided. In the same sense as stated above, lower layers mean less effort for overhead processing. In some cases it might be strictly necessary to use higher layers, if switching nodes of the respective layer reside within the protected domain. Otherwise the sequencing of data units on redundant paths might not be identical and synchronization becomes impossible. For example, if ATM-VP-crossconnects reside within the domain, the multiplexing of cells from different ATM-VP-connections on the respective outputs of two redundant switches may create different cell sequences, due to different internal states and buffers. Solely within the VPs the cell sequence is guaranteed. Therefore, in this example it would be necessary to implement the selection at the RDG for the ATM-VP-layer and process each VP separately.

Besides the overhead processing, the selection functionality consists of two main tasks, namely synchronization of the data streams and selection of one of them. To provide hitless recovery from failures a permanent synchronization of the redundant data streams is necessary. Different approaches may be identified to achieve this synchronization. Traditionally extra sequence information is added to the data stream, for example sequence numbers or synchronization cells. Besides the need for additional bandwidth, this also requires a structure of the data units that accepts this sequence information (for example the small header in ATM cells does not offer the possibility to add sequence information). Furthermore, special devices to insert this extra infor-

mation are necessary at the sending side, too. A more sophisticated way to achieve synchronization is to correlate the user data of the redundant data streams. So no extra bandwidth and no extra functionality at the sender have to be provided. The whole functionality is concentrated at the receiver. In [OhUe93] an algorithm for this purpose has been proposed. The algorithm has the drawbacks, that it takes rather long to recognize the synchronization and it does not tolerate the loss of single data units (cells). An improved algorithm has been developed and is currently being patented.

For the selection of one of the synchronized data streams, two different approaches exist. Selection of the leading data stream results in a minimal propagation delay, whereas selection of the lagging data stream allows the correction of loss of single data units, since data is not forwarded until it has been received from both data streams. Depending on the requirements one approach might be chosen.

Conventional recovery techniques use distributed error detection and signaling mechanisms (mostly OAM mechanisms, e.g. AIS, RDI) to trigger the protection switching operation. This leads to interruptions in the data stream due to time constants in the detection process and propagation delay of the signaling messages. In this paper it is proposed that the error detection is placed directly at the switching function. Corrupted data units are detected using CRC check and removed if they cannot be corrected. Loss of single data units might be tolerated or corrected. Lost data units are detected by the synchronization mechanism. Full breakdown of one link is assumed, if more than one data unit in direct sequence is missing or in error. An extension to tolerate or correct even more loss is conceivable. With this approach, it becomes possible to switch to the protection channel with zero loss.

As already mentioned before, the 1+1-redundancy technique for non-stop operation may be applied to different network layers. In Figure 6 an overview of the properties of the application on different layers is given. Obviously, the number of terminated OAM channels increases with the higher layers. Also increasing is the selectivity and the multiplexing factor, especially in the ATM layer. This leads to a high implementation effort in the higher layers, since every virtual connection has to be processed individually. Looking at today's available semiconductor technology, it should be possible to integrate the selection functionality for the layers up to the ATM-VP-layer in a single chip.

For the application of the new technique on existing networks and with existing equipment it is important to regard the interactions with common protection switching mechanisms. Obviously, no problem arises, if the new protection technique resides in a layer below the existing mechanism. For example a duplication-selection-redundancy domain might be applied to a SDH subnetwork without affecting an DRA (distributed restoration algorithm) in the VP layer. Conversely, if lower layers are protected using a conventional mechanism and higher layers (e.g. ATM-VP) use the redundancy domain approach, it might come to states where the two mechanisms influence each other disadvantageously. For example a redundancy domain in the VP layer might recover instantly from a crossconnect failure without loss, while in the SDH layer rerouting is started. Therefore, it would be necessary to coordinate the recovery techniques.

	AAL-Link	VC-Link	VP-Link	TP-Link	MS-Link	RS-Link	PS-Link
Terminated OAM Channels	F5	F4	F3	F2	F1	-	-
Selectivity	ATM-VC		ATM-VP	VC-4	none		
Multiplexing	1	max. 2^{16}	max. 2^{12}	max. 64	1		
Implementation Effort	very high	high	medium			low	very low
	No single chip solution possible		VLSI Single chip solution possible				
Interaction with existing recovery mechanisms	Interactions with SDH protection switching must be considered				Transparent for SDH protection switching		
	Interactions with ATM-VP- protection switching must be considered			Transparent for ATM-VP protection switching			

Figure 6 Properties of DS-RDGs in different network layers

3.3 Redundancy Domains for Diversity

Another approach to guarantee non-stop-operation in the case of a failure are diversity techniques. As already introduced, they reduce the required bandwidth by distributing the data stream on multiple paths and transmitting redundant information on supplementary paths (Figure 7).

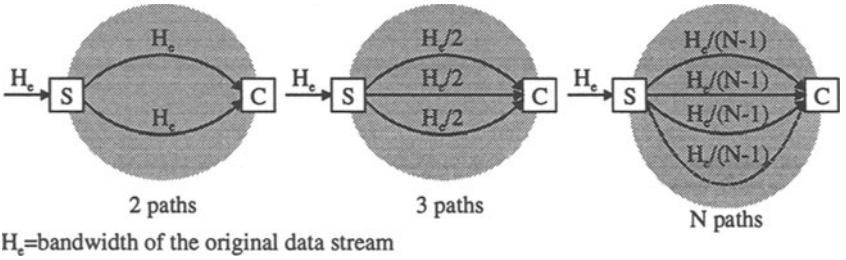


Figure 7 Protection with Diversity

Although the availability may be slightly reduced compared to 1+1-redundancy, a hitless recovery can be guaranteed for all recoverable errors. The required bandwidth B_{total} and the resulting availability A_{total} for this (N-1)-of-N system in dependence of the number N of redundant paths are

$$B_{total} = B_e \frac{N}{N-1} \quad , \qquad A_{total} = NA_S{}^{N-1} + (1-N)A_S{}^{n},$$

with B_e = bandwidth of the complete input path, A_s = availability of one single path.

Figure 8 shows the interdependence of unavailability and relative bandwidth for different numbers of redundant links. Obviously, already with a few paths, a lot of bandwidth can be saved compared to a full redundant transmission on two paths, while the reduction of the unavailability is less than one order of magnitude.

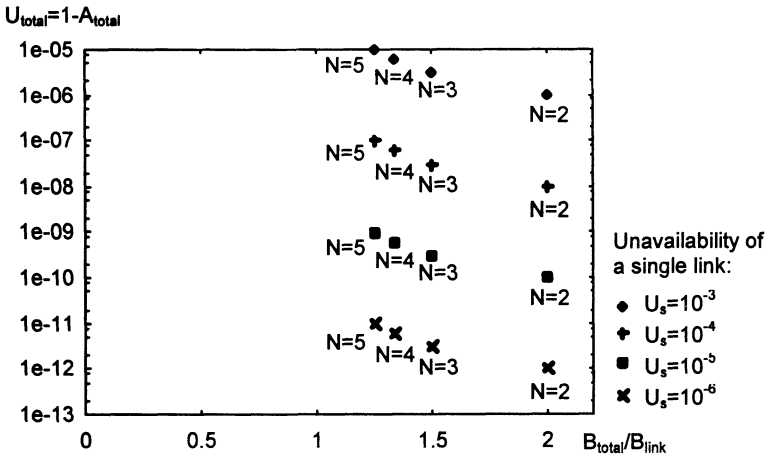


Figure 8 Interdependence of unavailability and relative bandwidth for different numbers of redundant links

For the application of the diversity technique to redundancy domains, the data streams have to be split and sent on several paths when entering the domain. Before leaving the domain the data streams from these paths have to be combined and the original data has to be reconstructed. The principle structure of the RDGs for diversity protection is very similar to that for DS-gateways. The protocol for the underlying layers has to be processed before and after the splitting or combination function.

Two main challenges have to be solved for redundancy approaches: coding/decoding and synchronization of multiple data streams. The coding operation should work very fast and it should be applied to short data units to minimize additional delay. Further, it must not be too complex, to simplify implementation and maintain a high availability of the coder and decoder. Distributing data units on the paths in a round robin fashion and transmitting parity information on an additional path is a simple approach. The second challenge is the synchronization of multiple data streams. With additional synchronization information, for example sequence numbers, a simple operation is possible. The effort for synchronization based on the user data, as described for the synchronization of two data streams, raises exponential with the number of parallel paths and is therefore hardly appropriate. With a combination of both approaches the advantages can be combined. Simplified synchronization detection using sequence numbers in the search phase and exact synchronization using user data correlation in the working phase. Such algorithms are currently investigated and developed.

3.4 Redundancy for RDGs

In general, the two parts of a gateway (e.g. duplication and selection) are combined in a single device. Figure 9a shows a DS-gateway. The selection element in this non-redundant simple DS-gateway selects one incoming data stream from the redundant

paths and transfers it to the duplication block, where it is again duplicated and sent towards two different outputs. Similarly in the SC-gateway (Figure 9b) first the incoming data streams are combined and then split to be sent on several separate paths.

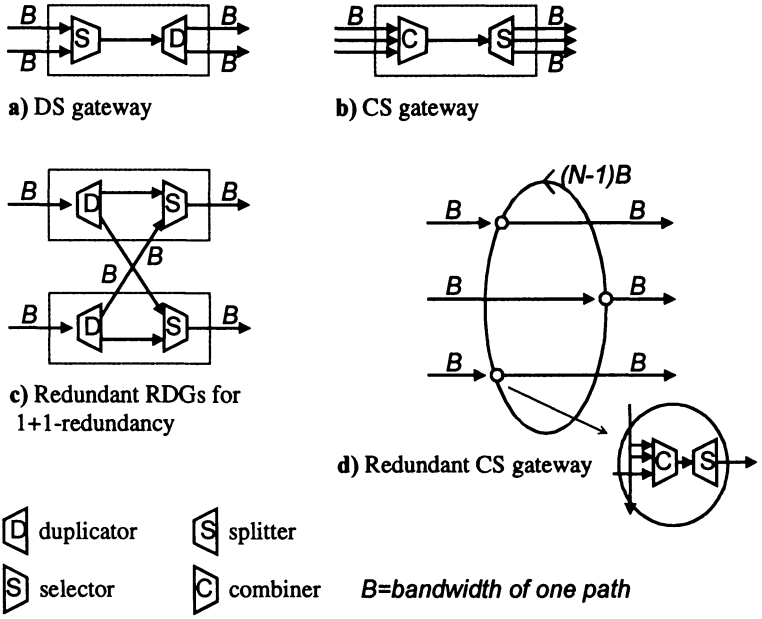


Figure 9 Redundancy for redundancy domain gateways

As already mentioned before, the fault tolerance of the RDGs itself plays a major role. Redundant provision of RDGs has already been sketched in Figure 4. A more detailed schematic is shown in Figure 9c. Two devices, each consisting of a duplicate and a select function, are connected via cross links. This allows to survive the failure of a single device or parts of it.

For the case of diversity, the redundancy for the gateways becomes more complex. One possible approach is shown in Figure 9d, where for each parallel path one RDG is foreseen. The gateways are logically fully meshed using a physical ring structure. Each gateway forwards one path. If the input path fails, the data can be reconstructed using the information of the other paths.

4 CONCLUSION

Redundancy domains are a new architectural concept for recovery in fault-tolerant networks, that allows to implement the recovery functionality outside the usual cross-connects or switches and thus to protect them externally instead of the provision of internal redundancy. Redundancy domains may cover any part of a network and may therefore not only be used for the external protection of single nodes but also for the protection of links or whole subnetworks independently of the physical location of crossconnect or switching systems.

To achieve non-stop-operation for connections in the case of node or link failures, two alternative approaches have been investigated concerning their application on these redundancy domains. While 1+1-redundancy is easy to implement and only requires two redundant paths, it uses a lot of bandwidth, whereas with diversity techniques the bandwidth requirements can be reduced, leading to a slightly increased unavailability..

Currently, our work is ongoing in the refinement of the coding and synchronization for diverse transmission. Further, availability measures for redundancy domains are investigated.

The work presented in this paper has been funded by Siemens AG, Munich. The author would like to thank Dr. K.-U. Stein and Dr. M-N. Huber for their helpful comments and valuable discussions.

References

[Ed95] Edmaier B.: Pfad-Ersatzschalteverfahren mit verteilter Steuerung für ATM-Netze, PhD thesis, Institute for communication networks, Technische Universität München, published at Herbert Utz Verlag Wissenschaft, München 1996, ISBN 3-89675-112-3

[FFGL91] Fischer W., Fundneider O., Göldner E.-H., Lutz K.A.: *A Scalable ATM Switching System Architecture*, IEEE Journal on Selected Areas in Communications, Vol. 9, Nr. 8, p.1299-1307, October 1991

[GuKa97] Gustafsson E., Karlsson G.: *A Literature Survey on Traffic Dispersion*, IEEE Network, p.28-36, March/April 1997

[G.805] ITU-T Specification G.805: Generic Functional Architecture of Transport Networks

[I.311] ITU-T Specification I.311: B-ISDN General Network Aspects

[I.610] ITU-T Specification I.610: B-ISDN Operation and Maintenance Principles and Functions

[It91] Arata Itoh: *A Fault-Tolerant Switching Network for B-ISDN*, IEEE Journal on Selected Areas in Communications, Vol. 9, Nr. 8, p. 1218-1266, October 1991

[Kr95] Krishnan P.: *An Efficient Architecture for Fault-Tolerant ATM-Switches*, IEEE/ACM Transactions on Networking, Vol. 3, No. 5, p. 527-537, October 1995

[KrDP95] Krishnan K.R., Doverspike R.D., Pack C.D.: *Improved Survivability with Multi–Layer Dynamic Routing*, IEEE Communications Magazine, July 1995

[OhUe93] Ohta H., Ueda H.: *Hitless Line Protection Switching Method for ATM Networks*, International Conference on Communications, Proceedings ICC '93, p.272-276, 1993

[St95] Stein K.: *Redundant cell or packet oriented Network*, application for a patent, Siemens AG, 1996

[YaSi91] Yang S.C., Silvester J.A.: *Reconfigurable Fault Tolerant Networks for Fast Packet Switching*, IEEE Transactions on reliability, Vol. 40, Nr. 4, p.474-478, 1991