Implementation of two Resilience Mechanisms using Multi Topology Routing and Stub Routers

Stein Gjessing
University of Oslo and Simula Research Laboratory
Oslo, Norway
Email: steing@ifi.uio.no

***Abstract*— Resilient Routing Layers (RRL) and Multiple Routing Configurations (MRC) have been proposed as methods to achieve fast recovery from router and link failures in connectionless networks. In this article we show how RRL and MRC can be implemented using the Multi Topology routing scheme and the Stub Router advertisements currently developed within the IETF. This makes RRL and MRC very viable candidates for protection of IP traffic in the next generation Internet.**

***Index Terms*— Network Resilience, Multi Topology Routing, Stub Routers, Resilient Routing Layers, Multiple Routing Configurations.**

I. Introduction

The Internet has become a commodity platform for business critical communication and real-time services; i.e. an infrastructure that more and more people rely on. The daily frequency of failures [1], however, highlights the need for fast, efficient and user transparent recovery. We have previously proposed two methods for very efficient and fast recovery called Resilient Routing Layers (RRL) [2] and Multiple Routing Configurations (MRC) [3]. Section II of this paper gives a short introduction to RRL and MRC.

RRL and MRC works at the network layer, and require modifications in both the the routing and the forwarding planes in order to handle link and node failures. We have noticed earlier that the routing layers of RRL and the routing configurations of MRC closely resembles the mechanisms that are contained in the Multi Topologies (MT) routing proposal for IP routing ([4] [5] [6]). Section III gives a short overview of MT.

We believe that RRL and MRC are very suitable and cost effective network layer protection methods. In this paper we show that both are implementable using mechanisms that are currently proposed within the IETF. Hence we believe that our resilience methods offer very viable solutions for protection of IP traffic in the future Internet.

This paper is organized as follows. After introducing RRL and MRC, we give a short overview of MT and how other researchers also have proposed to use MT for resilience. An additional concept we will use, also developed within the IETF, called Stub router advertisements [7], will also be introduced. Then it is shown how the implementation requirements of RRL and MRC can be met by MT and Stub router advertisements, and we also show in some detail what such an implementation must contain. Finally we conclude and point to further work.

II. Resilient Routing Layers and Multiple Routing Configurations

Resilient Routing Layers (RRL) is based on the idea of building spanning sub topologies over the full network topology, and using these sub topologies to forward traffic in the case of a failure. In RRL, we denote the sub topologies layers (as the name RRL indicates); in the sequel, however, we use the term *backup topology* instead. The backup topologies are built such that there exists a valid path between any pair of nodes in every backup topology. In RRL some links are removed from the original topology in order to create a backup topology [2]. In another approach, termed Multiple Routing Configurations (MRC), we assigned new weights to some of the links, thus creating backup topologies with all the same links, but with different (i.e. higher) costs on some of the links [3]. These higher costs links would usually be avoided by the forwarding process. In both RRL and MRC, links and nodes that are avoided by the forwarding engine are termed *isolated* network elements.

Figure 1 shows an original topology and three RRL backup topologies. The top part of the figure shows the full topology, also called the original topology, or topology 0. For each link and node, there exists at least one backup topology in which that network element is isolated. A router is isolated in a backup topology if it does not transit traffic. In figure 1 such nodes are dotted. A link is isolated if it does not have to carry any traffic.

In figure 1 the dashed links are isolated, and they can be considered to not be part of the (backup) topology.

If Figure 1 was to illustrate MRC, then the three backup topologies would be called configurations, as the name MRC indicates. The link weights on all the dashed links, as well as the link weights on all the solid line links attached to the isolated nodes, would be higher than their original cost. The rest of the solid line links would maintain their original cost.

Using a clever assignment of costs to links, we are able to show that configurations can be created that will always make it possible to route around a failed network element using the shortest path in one of the configurations [3]. In the sequel, when there is no ambiguity, the term backup topology will also be used for an MRC configuration.

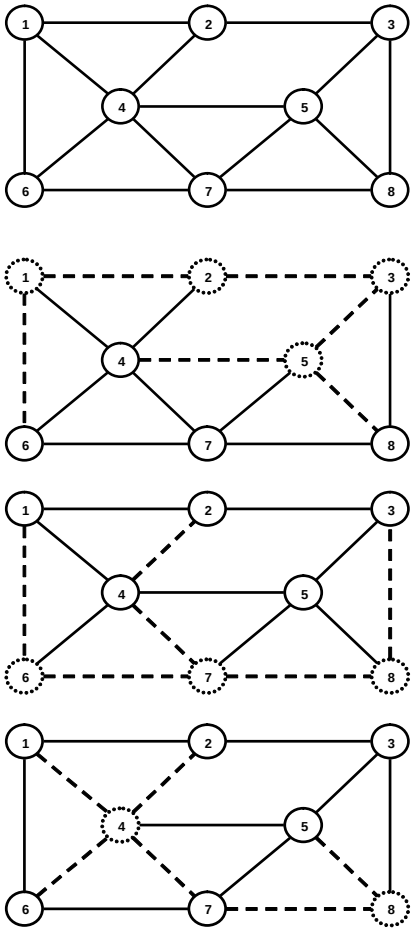

Fig. 1. The original topology at top, and three RRL backup topologies below. Notice that backup topologies are in general not (but may be) trees. Also note that a node or a link may be isolated in more than one backup topology.

A. Backup topologies and routing tables generation

Backup topologies can be generated by hand or automatically. We have developed several algorithms that generate backup topologies both for RRL and MRC [8] [9]. We have tested our algorithms on both real and synthetic topologies, both small and large. For network topologies with up to 512 nodes, it seems that at most 6 backup topologies are needed in order to isolate all links and all nodes.

We have not yet developed the full proof for the minimum number of backup topologies needed for a given original topology, but is seems that the length of the longest, pure ring or cycle part of the toplogy decides the number. E.g. in a pure ring topology with n nodes and n links; n/2 backup topologies are needed if one wants to isolate all nodes, and n backup topologies are needed to isolate both all nodes and all links.

In each backup topology, a shortest path algorithm is run in each node in order to generate a routing table for that backup topology. Hence, if there is one original and three backup topologies, conceptually, a node will have a total of four routing tables.

B. Forwarding in case of a failure

In the normal case, all traffic is forwarded in the original topology, i.e. according to the main routing table in all nodes (topology 0). When a packet arrives at a router, the forwarding engine must know which routing table to use. Conceptually this is done by marking all packets with a topology number that identifies the topology the packet is currently routed in. Initially all packets are marked so that they are routed in topology 0.

When a router discovers a link or a node failure, all packets that were supposed to be routed over the failed link or node, now has to change topology, i.e. they must from now on (and until they arrive at the egress node) be routed in one of the backup topologies. The chosen backup topology is one in which the failed element is isolated, and the router marks the packets with an identification of this chosen backup topology.

A packet that is routed in a backup topology may not revert back to be routed in the original topology, because that could in general create loops in the routing. To avoid loops, if a packet is routed in a topology other than the original, and encounters (another) failed network element, it may not freely change backup topology again. However, it is possible to change backup topologies more than once in a controlled way, e.g. by always moving to a higher numbered backup topology.

If the ingress node becomes aware of the failure, it has the option to change the topology of all packets that should have used this failed element, immediately. This will mean that the affected traffic will be routed in a

topology other than the original all the way from the ingress to the egress.

Notice that all traffic that is not affected by the failure is always routed in the original topology.

III. Multi Topology Routing

The main idea behind Multi Topology (MT) Routing is that different streams, different service levels or different services (e.g. multicast and unicast) may be routed and forwarded in different topology images. There have been three internet drafts available for the description of this, one for OSPF version 2, using IPv4 [4], one for OSPF version 3, using IPv6 [6], and one for IS-IS, also mainly addressing IPv6 [5].

In addition, RFC 3137, titled "OSPF Stub Router Advertisement" [7], describes a technique for isolating routers by creating what the authors call *stub* routers. A stub router will not carry any transit traffic, only act as an ingress or egress router for the network. The reason for a router to be a stub router could be that it has too few resources to handle its task as a transit router, e.g. because it is in the process of being introduced into, or removed from the network. The declaration of a router as a stub router could also be a general Traffic Engineering mechanism.

According to RFC 3137, a stub router is created by setting the cost of all but one link into the router to infinity. The remaining link keeps its original cost. This is very much the same as is done when creating backup topologies in RRL. In the sequel we will see that the stub router concept, in conjunction with Multi Topology Routing, may indeed be used to implement RRL.

A. Distribution of topology information

In MT-OPSFv2, it is suggested to use the TOS field of the LSA packets to advertise the link cost in the different topologies. The TOS field was originally created to advertise different costs for different DiffServ types of services. This use is however deprecated, so this field is now available.

In MT-OSPFv3, a Multi-topology bit is proposed in the Options field to signal that the originating router is MT capable. There is also a new type bit in the LSA, indicating an MT type LSA. An 8 bit field called MT-ID defines up to 256 different topologies. MT-ID 0 is reserved for LSAs that carry data about the original topology. A field called MT-ID Metric contains a metric associated to this MT-ID.

For each topology a Shortest Path First (SPF) algorithm is run, using link state information from that topology only.

B. MT Forwarding

The IETF drafts for Multi Topology Routing specifies that a packet is routed in one and the same topology from ingress to egress. The forwarding must be such that only routes belonging to that specific topology is used. However, when we use MT to implement resilience, all routes are contained in all topologies, so the last point will not be an issue.

All three MT drafts more or less leave it up to the implementation to decide how the forwarding of frames according to the different configurations should be performed. MT-OPSFv2 and MT-OSPFv3 both specifies that it is outside the scope of the specification how frames are forwarded in the topologies.

MT-IS-IS, however, suggests using different destination addresses for different topologies. This can of course be done for the OSPF versions too. Another method proposed for IS-IS is that if the DiffServ Code Point (DSCP) bits are not all used for QoS purposes, then some could be used as topology number.

C. Related work using MT for resilience

Menth and Martin [10] also propose to use MT for resilience. They assume that for each possible network element failure, there exists an entry in an MT forwarding table that forwards traffic by avoiding this failed element. This MT id is then stored in the packet header, and if no more erroneous network elements are encountered, the traffic is routed in this topology all the way to the destination. If, on the other hand, more failing network elemts are encountered, the traffic can change topology again. A time to live field in the packets prevents the traffic from looping forewer in the network. Menth and Martin, however, have no strategy when it comes to how the different topologies should be generated.

Sheffel et al. [11] also looks at MT Routing in order to see how many topologies are needed to optimally protect a network from link failures. They formulate a set of constraints and use binary linear integer programming to minimize the cost of routing in a backup topology image after the occurence of an error. They find that in a concrete network (the COST 239 network [12]) as few as 3 topologies are needed in order to perform optimal MT Routing after a link failure.

IV. Implementation

In this section we describe how RRL and MRC can be implemented using MT and Stub routers. We assume all routers in the network under consideration implement MT. Obviously the number of backup topologies in RRL or MRC is going to be the number of additional

topologies in MT. Hence MT-ID# n is the RRL or MRC backup topology number n. RRL and MRC default topology (topology 0) is also the default MT routing topology, i.e. MT-ID# 0. We assume that all routers participate in all topologies.

A. Building Routing Tables

We assume that RRL and MRC backup topologies are built centrally based on knowledge of the network topology. Each router is then informed about the backup topology(s) they are isolated in, and also the isolated backup topologies of each of its adjacent links and nodes.

When all routers have received this information they start to exchange Multi Topology Link State Advertisements (MT-LSA). Implementing RRL, the isolated routers use the Stub Router advertisements to tell the rest of the network that they are stub routers (i.e. that they are isolated), which links have infinte cost (the dashed links of figure 1) and also which links can still be used to reach the isolated routers. All routers have to advertise its state (and the state of its links) in all topologies (the original as well as all the backup ones).

When impelementing MRC, Stub Router advertisments are not used. Instead the links are advertised with the costs that were the result of the initial building of the MRC backup topologies. Hence most links have their original costs, while other links, in some topologies, are advertised with a new (and higher) cost (in order to be primarily avoided by the forwarding engine).

When a router has received all necessary link state information in all topologies, it builds one shortest path forwarding table for each topology. To implement such tables we have two choices. If each topology has its own address space (see below), the best thing is to build one table per topology. If, on the other hand, it is possible to mark each packet with a topology number, there should be one larger table whose entries are structures mapping topology numbers to out links or next hops.

B. Error detection and forwarding

It is stated in the IETF MT drafts that the same topology should be used to forward a packet all the way from ingress to egress. The reason for this is simply to avoid looping. Hence, if looping can be avoided by other means, this requirement can be relaxed.

All three MT drafts leave it up to the implementation to map the incoming packets to a topology, and hence choose the routing table to be used for forwarding. As is suggested for IS-IS, a separate address space for each topology could be used. In IPv6 the address space size is large, so this could be a viable solution.

Since we are going to implement RRL and MRC by letting packets change topology inside the network (see below), a mapping between destination addresses are then needed in each router. A function with three parameters must be implemented. The inputs to the function are the packet's old topology number, the old address and the new topology number. The function produces the packet's corresponding new address in the new topology.

The mapping between addresses in the different topologies could be created by the same system that generates the backup topologies, and should be distributed to all routers. An alternative would be to find a deterministic algorithm that all routers could use to calculate the mapping. Anyhow the mapping (or possibly only the algorithm) has to be stored in all routers.

A more intuitive and appealing method would be to mark the packet header with the topology number. In IPv6 this seems like a viable solution since the header is very flexible with the possibility to be extended. In IPv4 we may, as suggested by the MT-IS-IS draft, use some of the TOS bits if they are available. As pointed out before, in all the original topologies we have studied (both real and synthetic) we have never come across a network topology where more than 6 backup topologies are needed. Hence 3 bits are enough.

In RRL and MRC a packet is routed in the original topology until an error is encountered. When the out link or the next hop router (that is not the egress router) has failed, the packet is taken out of the regular forwarding path for special treatment. All routers also contain a function from a possibly failed network element (a connected link or a next hop router) to a set of backup topologies. This set contains the topologies in which this network element is isolated. Usually the incoming packet is routed in the default topology (topology number 0), but in general it may already have encountered one (or more) errors, so it might also be routed in a topology with a number higher than 0.

The error handling process first takes the packet out of the normal forwaring process. It then picks the backup topology with the lowest identification that is higher than the number in the incoming packet. If such a number is not found, the packet is discarded (in order to avoid loops). Notice that a packet will never be discarded on encountering its first error in the network. By allowing packets to change to topologies with higher numbers only, loops are avoided, and hence our error handling procedure will not conflict with the MT specifications.

This change of the packet's topology, as describe above, must be implemented in each router. In fact this change of the packet's topology when encountering an

error, is the main new mechanism in our implementation. When a packet has completed this change, it is put back into the regular forwarding path and forwarded according to its new topology identification.

V. Conclusions and Further Work

The author is part of a research team that has previously proposed two frameworks for network protection termed Resilien Routing Layers (RRL) and Multiple Routing Configurations (MRC). RRL is based on a set of backup topologies that are sub-topologies (some links are removed) of the complete network, while MRC is based on the original network topology with modified link costs. Multi Topology (MT) Routing is a new paradigm developed within the IETF, that can be used when several different views of the network are needed. IETF also introduces the concept of a Stub router, when traffic should avoid transiting a certain router.

This paper has discussed the implementation of our two IP protection methods, RRL and MRC, within the MT and Stub router proposals. It has been shown that such implementations are possible; indeed it turns out that MT is an ideal basis for the implementation of RRL and MRC. The Stub router definition closely matches the notion of isolated routers in RRL. Finally different implementations details were proposed and discussed, including how packets are handled when an error is encountered.

Others have also suggested to use MT for resilience, but they do not have any concrete proposals for the generation of backup topologies, nor how have they analysed in the same detail the performance of their proposed schemes, or given the same level of implementation details within the MT and Stub router proposals.

We have already made software that generates backup topologies. Our next step will be to make a prototype implementation of MT and Stub Router advertisements on routers in a small laboratory network, and then also carry out the implementation task described in this article. This will result in a test bed that can be used for the first assesments of the performance of our resilience methods in practise.

Acknowledgments

The author first of all wants to thank Audun Fosselie Hansen for many interesting discussions related to this article, and also for his help in preparing the manuscript. Thanks also to Amund Kvalbein for help in the reported research. Finally, thanks go to the rest of the team that develped RRL and MRC (Tarik Cicic and Olav Lysne).

References

[1] G. Iannaccone, C.-N. Chuah, R. Mortier, S. Bhattacharyya, and C. Diot, “Analysis of link failures in an ip backbone,” in *2nd ACM SIGCOMM Workshop on Internet Measurement*, Nov. 2002, pp. 237–242.

[2] A. F. Hansen, T. Čičić, S. Gjessing, A. Kvalbein, and O. Lysne, “Resilient routing layers for recovery in packet networks,” in *Proceedings of International Conference on Dependable Systems and Networks (DSN)*, June 2005.

[3] A. Kvalbein, A. F. Hansen, T. Čičić, S. Gjessing, and O. Lysne, “Fast IP network recovery using multiple routing configurations,” in *Proceedings of IEEE/INFOCOM*, Apr. 2006.

[4] P. Psenak, S. Mirtorabi, A. Roy, L. Nguen, and P. Pillay-Esnault, “MT-OSPF: Multi topology (MT) routing in OSPF,” IETF Internet Draft, Apr. 2005, draft-ietf-ospf-mt-04.txt.

[5] T. Przygienda, N. Shen, and N. Sheth, “M-ISIS: Multi topology (MT) routing in IS-IS,” Internet Draft, May 2005, draft-ietf-isis-wg-multi-topology-10.txt.

[6] S. Mirtorabi and A. Roy, “Multi-topology routing in OSPFv3 (MT-OSPFv3),” Internet Draft, 2005, draft-ietf-ospf-mt-ospfv3-00.txt.

[7] A. Retana *et al.*, “Ospf stub router advertisement,” in *IETF*, RFC 3137, June 2001.

[8] A. F. Hansen, A. Kvalbein, T. Cicic, S. Gjessing, and O. Lysne, “A comparison of different approaches for calculating resilient routing layers and multiple routing configurations,” Simula Research Laboratory, Technical Report 15-2005, 2005. [Online]. Available: http://www.simula.no/departments/networks/.artifacts/ RRLcomparison

[9] A. Kvalbein, A. F. Hansen, T. Čičić, S. Gjessing, and O. Lysne, “Fast recovery from link failures using resilient routing layers,” in *Proceedings 10th IEEE Symposium on Computers and Communications (ISCC)*, June 2005.

[10] M. Menth and R. Martin, “Network resilience through multi-topology routing,” University of Wurzburg, Institute of Computer Science, Tech. Rep. 335, May 2004.

[11] M. Scheffel, C. Gruber, T. Schwabe, and R. Prinz, “Optimal multi-topology routing for IP resilience,” *To appear in AEÜ International Journal of Electronics and Communications*, 2006.

[12] M. J. O’Mahony, “Results from the COST 239 project. Ultra-high capacity optical transmission networks,” in *Proceedings of the 22nd European Conference on Optical Communication (ECOC’96)*, Sept. 1996, pp. 11–14.