A Comparison between Two Post-Failure Load Distribution Techniques with Multiple Routing Configurations

Mohamed A El-Serafy, Moustafa H Aly
Arab Academy for Science, Technology and Maritime Transport (AASTMT)
Alexandria, Egypt
mohamed.elserafy@aast.edu, drmosaly@gmail.com

El-Sayed A El-Badawy, Ibrahim A Ghaleb
Electrical Engineering Department
Faculty of Engineering, Alexandria University
Alexandria, Egypt
sbadawy@hotmail.com, ibrahim_ghaleb@yahoo.com

Abstract: When a link fails in an IP network, the network re-converges to redirect the traffic away from the failed link. IP fast reroute is used to minimize the time taken to redirect the traffic to an alternate path. Multiple routing configurations (MRC) is the IP fast reroute scheme under consideration. In this paper, we will discuss how to minimize the impact of the MRC recovery process on the post failure load distribution over network links. Two techniques used to achieve a good load distribution across links after failure are compared. The first technique utilizes manual link weight manipulation with MRC. The second technique is the modified MRC. Both advantages and disadvantages of the two techniques are discussed.

Keywords: Multiple routing configurations, IP fast reroute, multi topology routing, network protection, network utilization, network reliability.

I. INTRODUCTION

The role of the Internet is continuously increasing and many technical, commercial and business transactions are carried out by millions of users who exploit a set of network applications [1]. An increasing fraction of computing and storage is migrating to a planetary cloud of datacenters [2]. Cloud computing has emerged as a popular computing environment that provides computing power and storing space for small to large enterprises, within minimum cost [3]. The cloud consists of geographically distributed mega datacenters connected by a high capacity network [4].

In such a network, different services are replicated over multiple datacenters, so that a user request can be served by any datacenter that supports the specified service [5]. To meet the demands set by the high volume of traffic between datacenters, optical networks are ideally suited, given their high bandwidth and low-latency characteristics [5]. In such networks, path protection against network failures is crucial. A link or a node failure is typically followed by a period of routing instability. During this period, packets may be dropped due to invalid routes and this has an adverse effect on real time applications [6]. Traditionally, such disruptions have lasted for periods of at least several seconds. They have been shown to occur frequently and are often triggered by external routing protocols [7].

Recent advances in routers have reduced this interval to under a second for carefully configured networks using link state interior gateway protocols (IGP). However, new Internet services that are classified as real time applications are very sensitive to periods of traffic loss [8].

Addressing these issues is difficult because the distributed nature of the network imposes an intrinsic limit on the minimum convergence time that can be achieved. However, there is an alternate approach, which is to compute backup routes that allow the failure to be repaired locally by the router detecting the failure without the immediate need to inform other routers of the failure. In this case, the disruption time can be limited to the small time taken to detect the adjacent failure and invoke the backup routes. This approach is called IP fast reroute [9]. Link state IGP uses re-convergence process that is characterized by being reactive and global. IP fast reroute is proactive and local in nature. One of the IP fast reroute approaches is MRC [10], MRC was built on the concept of resilient routing layers (RRL) [11] and [12]. In this paper MRC is used as the main approach for IP fast reroute.

In this paper, we compare between two techniques that are used to minimize the impact of the MRC recovery process on the load distribution for traffic across links after failure. It is shown that the second technique is better than the first one, however it still has some drawbacks. We conclude that this problem needs a new approach to achieve a good load distribution after failure that suits any network topology with any traffic demand matrix. The rest of this paper is organized as follows: in Section 2, the idea behind MRC is outlined. In Section 3, post failure load distribution with manual link weight manipulation using our own simulation results is presented. In Section 4, the modified MRC is outlined and finally findings are concluded in Section 5.

II. IP FAST REROUTE USING MRC

MRC is a proactive and local IP fast reroute scheme which allows recovery in the range of milliseconds. MRC allows packet forwarding to continue over preconfigured alternative next hops immediately after the detection of the
failure. Using MRC as a first line of defense against network failures, the normal IP re-convergence process can be put on hold. This process is then initiated only as a consequence of non-transient failures [10]. MRC guarantees recovery from any single link or node failure, which constitutes a large majority of the failures experienced in a network [13].

The main idea of MRC is to use the network graph and the associated link weights to produce a small set of backup network configurations. The link weights in these backup configurations are manipulated so that the node that detects the failure can safely forward the incoming packets towards the destination. This is achieved for each link and node failure, and regardless of whether it is a link or a node failure. MRC assumes that, the network uses shortest path routing like open shortest path first (OSPF) and destination based hop-by-hop forwarding [10]. In a configuration that is resistant to the failure of a node \( n \) or link \( l \), link weights are assigned so that traffic routed according to this configuration is never routed through node \( n \) or over link \( l \). In MRC, node \( n \) and link \( l \) are called isolated in a configuration, when no traffic is routed through \( n \) or \( l \) according to this configuration [10].

After generating the backup configurations, a standard link state routing protocol like OSPF is run over the generated configurations to construct a loop free configuration specific forwarding table to all destinations in the network. When a packet is forwarded according to a configuration, it is forwarded using the forwarding table calculated based on that configuration. Figure 1 shows packet forwarding process from a node’s perspective [10].

When a router detects that a neighbor can no longer be reached through one of its interfaces, it does not immediately inform the rest of the network about the connectivity failure. Instead, packets that would normally be forwarded over the failed interface are marked as belonging to a backup configuration, and forwarded on an alternative interface towards its destination. The packets must be marked with a configuration identifier. So, the routers along the path know which configuration to use [10].

MRC does not affect the failure free original routing. When there is no failure, all packets are forwarded according to the original configuration, where all link weights are normal. Upon detection of a failure, only traffic reaching the failure will switch configuration. All other traffic is forwarded according to the original configuration as normal [10].

Kvalbein et al. implemented the MRC algorithm using a Java program. The network used in the simulation was the European COST239 network. A detailed explanation of the MRC algorithm can be found in [10].

III. MANUAL LINK WEIGHT MANIPULATION FOR POST FAILURE LOAD DISTRIBUTION WITH MRC

Once a node running MRC detects a failure towards the final destination, there exists a backup configuration that will forward the traffic to its final destination on a path that avoids the failed element. This shifting of traffic to alternate links after a failure can lead to congestion and packet loss in parts of the network [14]. In this section, our simulation model using OPNET modeler software shows the impact of the MRC recovery process on the link load. The link load is shown to increase after single node failure. To solve this problem, and to achieve good load distribution across links, manual link weight manipulation is applied on the MRC backup configurations. The following is considered in our simulation model:

- The European COST239 network C0 shown in figure 2 [15].
- All link weights are normalized to 10 in C0.
- The traffic demand matrix consists of full mesh traffic between all topology nodes.
- IP fast reroute technique is MRC.
- Number of MRC backup configurations \( n = 5 \) (C1, C2, C3, C4, C5).
- In C1 isolated nodes are R0 & R6 and isolated links are given very high weights 10,000 and 30,000.
- OSPF is the IGP routing protocol.

![Figure 1. Flow chart of a node’s packet forwarding process using MRC [10]](image-url)
Let’s consider the following scenario: R1 is sending traffic to R9. There exist two equal weight paths from R1 to R9 in the forwarding table of R1, through R0 and R8. R1 will load balance the traffic over the two paths.

At the same time, R9 is sending traffic to R4, from R9 perspective there exist four equal weight paths to R4, through R0, R8, R6 and R10. R9 will load balance the traffic over the four available paths.

Considering the case of single node failure and let it be the failure of R0. R1 will detect the failure and instead of dropping the traffic that is destined to R9 that passes through R0, R1 will check the forwarding table of the backup configuration C1 (the backup configuration that has R0 isolated). In this backup configuration, R1 has just one best path to R9 through R8.

At the same time, R9 will detect the failure of R0 and will redirect the traffic that was destined to R4 and passing through R0 based on the backup configuration C1. Instead of having four equal weight paths, only two equal weight paths are available through R8 and R10.

As a consequence of R0 failure and redistributing traffic to other available links based on the backup configuration C1, the network starts to face congestion and some of its links are having link utilization above 50% of its capacity.

Figure 3 shows the backup configuration C1 under consideration by this example.

To decrease the load on the congested links, one can manually manipulate the link weights so as to get a better load distribution for the traffic.

To decrease the load on the link from R1 to R8, link weights need to be manipulated so as to find another path for the traffic from R1 to R9. Link weights on both R1-R8 link and R1-R4 link in the backup configuration C1 will be manipulated to get a better load distribution of the traffic from R1 to R9. Now, two equal weight paths are available for R1 when sending traffic to R9.

As a consequence to manually manipulating link weights, the traffic from R9 to R4 will have only one best path in the backup configuration C1 instead of two paths.

The links utilization is analyzed to try to better understand what really happened when links weight are manually manipulated. The below three different scenarios are used as a benchmark for comparison:

- Failure free link utilization.
- Post failure link utilization.
- Post failure with load balance link utilization.

R1-R8 link utilization is analyzed as follows:

- Failure free link utilization is 460 Mbps.
- Post failure link utilization is 670 Mbps.
- Post failure with load balance link utilization is 150 Mbps.

Using the same criteria, R9-R10 link utilization is 390 Mbps, 580 Mbps and 460 Mbps.

It is clear that, link utilization increases after failure to above 50% of link capacity proving the impact of the MRC recovery process and then drops down after manually manipulating links weight. Figure 4 shows the above mentioned scenarios for both R1-R8 and R9-R10 links.
Similarly, the link utilization for R1-R4 link is 400 Mbps, 510 Mbps and 850 Mbps. The link utilization for R9-R8 link is 280 Mbps, 580 Mbps and 650 Mbps. Finally, the link utilization for R8-R1 link is 490 Mbps, 690 Mbps and 880 Mbps. The above three links have severe high link utilization which is a consequence of manually manipulating link weights to load balance the traffic from R1 to R9 on two equal weight paths.

As a drawback of applying manual link weight manipulation, the traffic from R9 to R4 used to have 4 equal weight paths in the failure free case and now it has only one path after. Figure 5 shows the above three links utilization.

As a conclusion, using manual link weight manipulation technique to solve the problem of achieving good load distribution in the network has the following advantages:

- Simple - no complex algorithm is needed.
- Easy to deploy in small networks.
- Used to achieve good load distribution for selected number of links.

Its disadvantages are as follows:

- Manual – Hard to deploy in large networks
- Cannot be used to achieve global load distribution for all network links at the same time.

IV. MODIFIED MRC FOR POST FAILURE LOAD DISTRIBUTION

As shown in the previous section, the effect of a failure on the load distribution when MRC is used is highly variable. Occasionally, the load added on a link can be significant as shown in figure 6 [16].

In this section, it is shown how Kvalbein et al. tried to solve this problem. Kvalbein et al. proposed an approach for minimizing the impact of the MRC recovery process on the post failure load distribution. They discussed how a good load distribution can be achieved in the network immediately after a link failure when MRC is used as a fast recovery mechanism.

They presented an algorithm to create the MRC backup configurations in a way that takes the traffic distribution into account. They referred to this new algorithm as the modified MRC. They presented a heuristic aimed at finding a set of link weights for each backup configuration that distributes the load well in the network after any single link failure. The scheme is strictly proactive, no link weights need to be changed after the discovery of a failure [17].

The load distribution in the network after a link failure depends on three factors [17]:

1. The link weight assignment used in the normal configuration C0.
2. The structure of the backup configurations, i.e. which links and nodes are isolated in each backup configuration C1 to Cn.
3. The link weight assignments used in the backup configurations C1 to Cn.

The link weights in the normal configuration are important since MRC uses backup configurations only for the traffic affected by the failure, and all non-affected traffic is distributed according to them. The backup configuration structure dictates which links can be used in the recovery paths for each failure. The backup configuration link weight
assignments determine which among the available backup paths are actually used [17].

Network operators often plan and configure their network based on an estimate of the traffic demands from each ingress node to each egress node. The knowledge of such demand matrix \( D \) provides the opportunity to construct the backup configurations in a way that gives a better load balancing and avoids congestion after a failure.

To do this, the modified MRC was proposed to construct a complete set of valid configurations in three phases as follows [17]:

1. The link weights in the normal configuration are optimized for the given demand matrix \( D \) while only taking the failure free situation into account.
2. The load distribution in the failure free case is considered to construct the MRC backup configurations in an intelligent manner.
3. The link weights are optimized in the backup configurations to get a good load distribution after any link failure, i.e., the recovered traffic is distributed over less utilized links.

The method for link weight setting is based on perturbing link weights using a local search heuristic. Kvalbein et al. adopted a modified version of the local search heuristic presented in [18]. The link weights in the backup configurations are optimized to give a good load distribution after any link failure.

Optimizing for all possible link failures does not scale well as network size increases, because of the number of evaluations needed. To overcome this problem, they assumed that only a few link failures are critical with respect to the load distribution after failure. They optimize the link weights taking only the critical link failures into account as presented in [19]. Since the optimization procedure uses random search and relies on an estimate of the traffic matrix, it is preferred to be implemented by a central network management unit. Since the optimization is a computing intensive process, it should be conducted as a background process [17].

The input to the algorithm for generating backup configurations is the network topology \( G \) and the number of backup configurations \( n \) to be created. The outcome of this algorithm is dependent on the network topology \( G \) and the traffic demand matrix \( D \). In order to limit the calculation complexity only link failures were considered in the optimization phase.

For small networks like COST239, link weight optimization is performed in seconds or minutes. For larger networks the optimizations can take several hours, and should be conducted as a background process. A more detailed description of this algorithm can be found in [16].

Figure 7 shows the worst case load (maximum load) on all unidirectional links in the COST239 network in the failure free case, after IGP re-convergence and when modified MRC is used to recover traffic. The links are sorted by the load in the failure-free case.

It is clear that, the worst case load peaks are reduced for the optimized MRC compared to the standard MRC shown in figure 6. The maximum link load after link failure has been reduced from 118% to 91%, which is better than what is achieved after a full IGP re-convergence. In figure 8 the standard MRC and the modified MRC are directly compared. The links are sorted by their maximum load after failure using standard MRC. The optimized MRC often manages to route traffic over less utilized links after the failure of a heavily loaded link [16].

The modified MRC algorithm has the following advantages:

- Automatic - no link weight manipulation is needed after the discovery of a failure.
- Used to achieve good load distribution across all links.

Its disadvantages are as follows:

- Complex - the complexity of the algorithm is affected by the size of the network topology \( G \) under consideration.
• A predefined demand matrix $D$ must be defined. If the matrix is changed, the algorithm has to run again.
• Cannot be used to protect against all link failure. The algorithm is designed to protect against the failure of critical links only by distributing the traffic that was held by these links on other available links.

V. CONCLUSION

In this paper, we compared between two techniques used to achieve good load distribution of traffic across network links after failure while using MRC as an IP fast reroute scheme. The advantages and disadvantages of both techniques were highlighted. Although the first technique is simple, it is clear that it is not scalable for large sized networks. It suits small sized networks where good load distribution is required for selected links. The second technique can achieve good load distribution across all network links, but it is complex and its complexity increases as the network size increases. Also, it is not designed to account for the failure of all topology links, instead it focuses on the failure of critical links. We conclude that a newer technique is needed that overcome the above mentioned drawbacks.

REFERENCES