A New Technique for Full Fast Recovery in Hop-by-Hop Routing

S Sae Lor, R Landa, and M Rio

University College London

Abstract: This paper proposes a novel technique for fast re-route in hop-by-hop routing in case of single link failures. It offers full repair coverage without requiring additional mechanisms such as tunnelling or interface-specific operations. The technique handles the failures without jeopardising the operable parts of a network. That is, the impact on network traffic is kept to minimum. To guarantee a 100% protection, a counting mechanism which indicates the number of repetitions the re-routed packets need to be forwarded along the alternate next hops is required. In addition, various metrics and topologies are used to evaluate the technique to ensure realistic results.

1 Introduction

Network reliability has become one of the most concern problems as it determines the deployability of emerging services and applications. More precisely, a wide range of upcoming services requires a network that is more resilient than the existing ones as a single packet loss can potentially lead to a severe damage to an organisation. Examples of such services are remote surgery and military applications. Thus, it is important to minimise the routing disruptions caused by failures.

Several approaches have been proposed in the past decades to alleviate the packet losses during network convergence. These include the use of Multi-Topology (MT) routing [1, 2], multi-path routing [3, 4, 5], IP Fast Re-Route (IPFRR) [6], and a specific type of an overlay network [7, 8]. Nevertheless, none of these techniques can provide an optimal solution to the network reliability problem. This paper proposes a new technique based on IPFRR approach to prevent the forwarding discontinuity due to convergence process.

Following this section, IPFRR framework and its well-known techniques are concisely described. After that, in Section 3, a novel technique known as fast re-route using Alternate Next Hop Counters (ANHC) is presented in details. In order to ensure practicality, an extensive evaluation is made in Section 4. Finally, this paper is concluded in Section 5.

2. IP Fast Re-Route (IPFRR)

An IPFRR framework is specified in the IETF Internet draft [6] for handling packet losses during network failures. In general, when a failure occurs, packets are dropped for a certain duration. This period is divided into the time taken to detect the failure, the time taken to react to the failure, the time taken to notify all routers in the network of the failure, and the time taken to implement new paths for affected destinations into the hardware. Thus, IPFRR requires two sets of mechanisms: *a)* mechanisms for fast failure detection and *b)* mechanisms for repair paths.

Recently, solutions for fast failure detection and fast convergence have been proposed [9, 10]. The techniques tweak the protocol parameters, *i.e.* Hello interval in OSPF and IS-IS, in order to detect the failure faster. This implies a shorter convergence time, which can potentially reduce the amount of packet losses. In contrast, creating backup paths is not as simple as adjusting the Hello interval, as traditional IP routing does not employ source-based routing. Consequently, several algorithms used to create backup paths may lead to routing inconsistencies. However, various IPFRR techniques that guarantee a loop-free environment have been proposed. These include Loop-Free Alternates (LFAs) [11], U-turn [12], tunnels [13], not-via addresses [14], and Failure Insensitive Routing (FIR) [15]. LFAs are the simplest fast re-route technique, yet its repair coverage heavily depends on the underlying network topology. U-turn and tunnels offer higher percentage of destinations being protected than LFAs. In contrast, not-via addresses and FIR guarantee a full protection for any single failures (only link failures for FIR). Regardless of the reliability these mechanisms provide, using not-via addresses may degrade the performance of a router due to the need for encapsulation/decapsulation. On the other hand, FIR requires interface-specific forwarding table to ensure loop-free re-routing and therefore, it is more difficult to implement in real hardware. Although these techniques may have different pitfalls, it is important to note that, IPFRR is a viable option for failure recovery in IP networks due to its minimal requirements, considerably wide repair coverage, and immediate reaction to failures.

3. Fast Re-Route Using Alternate Next Hop Counters (ANHC)

The routing technique using ANHC requires two algorithms to compute local alternate next hops (*i.e.* backup next hops) and the corresponding ANHC values.

3.1 Computing Alternate Next Hops

Basically, a failure-detecting router can easily re-calculates the new shortest paths for affected destinations by excluding the failed link from its database. However, this may lead to routing inconsistencies as the failure information is not globally known. Under fast re-route using ANHC, a router pre-computes the alternate next hop for each destination based on edge disjoint paths paradigm. This creates correlation between paths of all origins towards the same destination. Thus, a consistent routing can be performed using a mechanism called the "alternate next hop counting".

It must be noted that, edge disjoint paths do not always exist and hence, fast re-route using ANHC employs partially edge disjoint paths instead. The algorithm used to find the alternate next hops is simple. First the sum of all link weights in a network is denoted as W_t. To compute a maximally edge disjoint path, the weights of all links employed in the normal shortest path are increased by W_t. Then, based on this re-calculated weights, a well-known shortest path algorithm such as Dijkstra's algorithm is performed to obtain the alternate next hop.

3.2 Computing ANHC Values

In the presence of failures, re-routing packets based only on local alternate next hops cannot guarantee a loop-free forwarding. Thus, it is important to employ an alternate next hop counting mechanism which enables a loop-free environment. In general, when a packet encounters a failure, it will be marked with an ANHC value based on the destination. Each intermediate router determines this value whether it is non-zero. If not, it decreases the ANHC value by 1 and forwards the packet to its alternate next hop. After the ANHC value reaches 0, the packet can be forwarded along the normal shortest path without traversing the failed link.

Computing ANHC value requires two parameters: a) the alternate next hops of all origins to the same destination and b) the alternate path of the failure-detecting node. This counting mechanism ensures that the packet will traverse outside its normal branch in the shortest path tree rooted at the destination. That is, it will not encounter the same failure again if an alternate path exists.

First, the ANHC value is set to 0. The router compares the hops sequence of its alternate path with the local alternate next hops of the intermediate routers. For each comparison, if the hop sequence of the alternate path is equal to the local alternate next hop, the ANHC value is incremented by 1. Otherwise, the algorithm terminates and the ANHC value is stored for the corresponding node pair.

3.3 Packet Forwarding

In normal scenario, packets are forwarded along the optimal shortest paths. However, in the presence of failures, the failure-detecting node marks the packet with a single bit to indicate that it has already encountered a failure. In addition, it set the ANHC value in the packet header (TOS field and Traffic Class field are recommended for IPv4 and IPv6, respectively). Each router receiving a re-routed packet (including the failure-detecting router) decreases the ANHC value by 1 and forwards it to the local alternate next hop. This process iterates until either the packet encounters another failure and dropped, reaches the destination, or its ANHC value becomes 0 in which it will be forwarded along the normal shortest path.

4. Performance Evaluation

This section evaluates the performance of fast re-route using ANHC. The simulation is based on a self-implemented Java software. It is important to note that a packet-level simulator such as ns-2 is not necessary as the results can be directly obtained from the characteristics of a network topology.

The simulation is run on a machine with 2.16 GHz Intel Core 2 Duo processor and 2 GB memory. Due to space limitation, only Abilene [16] and Sprint inferred from Rocketfuel data [17] are used for evaluation. While real traffic matrices for Abilene is publicly available, a Breadth-First Search algorithm [18] and a simple gravity model [19] is used to generate realistic link capacities and traffic matrices for Sprint topology. It is important to note that, gravity model works only if the momentum of nodes are known. Therefore, for traffic analysis, a PoP-level of Sprint topology is used instead. In addition, the traffic matrices are scaled to saturate the link capacity so that the link utilisation never exceeds 1 after network re-convergence.

4.1 Overheads

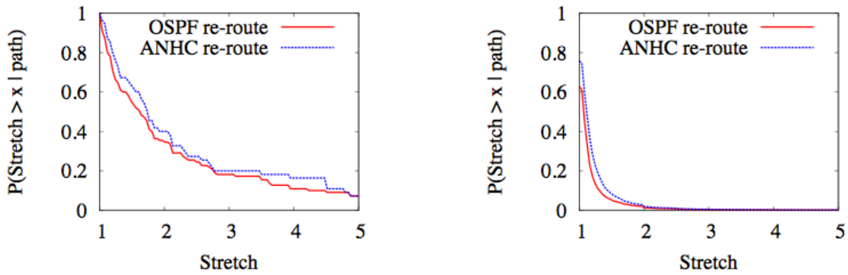
In this paper, three different overheads are evaluated: a) computational overhead; b) memory overhead; and c) packet overhead. The total computational overheads (alternate next hops and ANHC values computation time) are 0.114 ms for Abilene and 77.142 ms for Sprint. It must be noted that, Sprint consists of 315 nodes and 1944 links and hence, requires longer computation time than Abilene which has only 11 nodes and 28 links.

Fast re-route using ANHC does not require any additional entry to the routing table. However, it needs to enhance the information on the alternate next hop and ANHC value for each existing destination. This is considered minimal in exchange for higher reliability. From the simulation results, the optimal number of bits required in the packet header is 4 (1 bit is used to indicate a re-routed packet).

4.2 Routing Results

Two key properties of fast re-route using ANHC are full repair coverage and loop-free routing. Due to space limitation, the proofs are omitted in this paper, but a concise explanation is given. The repair coverage of the fast re-route using ANHC is 100% as the algorithm used to compute the alternate next hops avoids the local link failure. In addition, it guarantees a loop-free environment. This property is common due to the paths correlation between all origins to the same destination. That is, the ANHC never reaches 0 if the re-routed packet does not egress its original branch in the shortest path tree rooted at the destination. Once it arrives on a different branch (any other than the original), it will be forwarded correctly without creating a loop.

However, the path length stretch should be evaluated to ensure that the re-routing scheme does not incur very long alternate paths. Path length stretch is defined as the ratio between the cost of the alternate path and the normal shortest path. Figure 1 illustrates the path length stretch of the Abilene and Sprint networks.



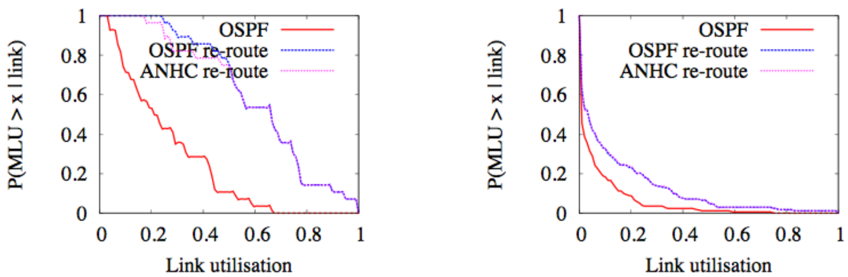
(a) (b)

Figure 1: Path length stretch of (a) Abilene and (b) Sprint.

It can be seen that the path length stretch of alternate paths provided by fast re-route using ANHC is very similar to that of OSPF re-route. Note that, the performance of fast re-route using ANHC is immediate while OSPF re-route has to wait until the convergence process completes.

4.3 Traffic Results

The impact of the routing technique on the network traffic can be evaluated using the Maximum Link Utilisation (MLU). Figure 2 illustrates the MLU of the Abilene and Sprint networks.



(a) (b)

Figure 2: Maximum Link Utilisation of (a) Abilene and (b) Sprint.

It can be seen that the MLU of both Abilene and Sprint networks is near optimal compared to the OSPF re-route.

5. Conclusions

The current routing architecture cannot accommodate emerging services and applications due to packet drops during convergence process. Several approaches have been proposed to alleviate the network reliability problem. However, none can guarantee an optimal solution. This paper proposed a novel technique based on IPFRR, which is a viable option. The technique employs a specific algorithm to pre-compute the alternate next hops used in conjunction with alternate next hop counting mechanisms.

It was proved that the routing strategy guarantees a full protection for any single link failures and a loop-free environment. The technique does not require any significant modifications to the traditional IP routing and has minimal impact on router's processing and memory. Results from the simulation showed that the recovery paths provided by fast re-route using ANHC is near optimal using OSPF re-route as a benchmark. Furthermore, the impact on traffic in term of the MTU is kept to minimal for practical topologies.

Acknowledgements

The authors would like to acknowledge their colleagues at NSRL, E&EE, UCL for their supports and recommendations. Also, the authors would like to thank Professor Mark Handley of the CS, UCL for his valuable comments on this work.

References

[1] A. Kvalbein, A. F. Hansen, T. Cicic, S. Gjessing, and O. Lysen. Fast IP network recovery using Multiple Routing Configurations. In *Proceedings of the 29th IEEE Conference on Computer Communications*, pages 23–29, Barcelona, Spain, Apr 2006.

[2] T. Cicic, A. F. Hansen, A. Kvalbein, M. Hartmann, R. Martin, and M. Menth. Relaxed Multiple Routing Configurations for IP fast reroute. In *Proceedings of the 10th IEEE/IFIP Network Operations and Management Symposium*, pages 457–464, Bahia, Brazil, Apr 2008.

[3] X. Yang and D. Wetherall. Source selectable path diversity via routing deflections. In *Proceedings of the ACM Special Interest Group on Data Communications Conference*, pages 159–170, Pisa, Italy, Sep 2006.

[4] W. Xu and J. Rexford. MIRO: Multi-path Interdomain ROuting. In *Proceedings of the ACM Special Interest Group on Data Communications Conference*, pages 171–182, Pisa, Italy, Sep 2006.

[5] M. Motiwala, M. Elmore, N. Feamster, and S. Vempala. Path splicing. In *Proceedings of the ACM Special Interest Group on Data Communications Conference*, pages 27–38, Seattle, WA, Aug 2008.

[6] M. Shand and S. Bryant. IP fast reroute framework. IETF Internet draft, Jun 2009. http://tools.ietf.org/html/draft-ietf-rtgwg-ipfrr-framework-11.

[7] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris. Resilient Overlay Network. In *Proceedings of the ACM Symposium on Operating Systems Principles*, pages 131–145, Banff, Canada, Oct 2001.

[8] D. G. Andersen, A. C. Snoeren, and H. Balakrishnan. Best-path vs. multi-path overlay routing. In *Proceedings of ACM Internet Measurement Conference*, pages 91–100, Miami, FL, Oct 2003.

[9] C. Alaettingoglu, V. Jacobson, and H. Yu. Towards milli-second IGP convergence. IETF Internet draft, Nov 2000. http://tools.ietf.org/html/draft-alaettinoglu-isis-convergence-00.

[10] M. R. Goyal and K. K. W. Feng. Achieving faster failure detection in OSPF networks. In *Proceedings of the IEEE International Conference on Communications*, pages 296–300, Anchorage, AK, May 2003.

[11] A. Atlas and A. Zinin. Basic specification for IP fast reroute: Loop-Free Alternates. RFC 5286, Sep 2008. http://tools.ietf.org/html/rfc5286.

[12]A. Atlas. U-turn alternates for IP/LDP fast reroute. IETF Internet draft, Feb 2006. http://tools.ietf.org/html/draft-atlas-ip-local-protect-uturn-03.

[13] S. Bryant, C. Filsfils, S. Previdi, and M. Shand. IP fast reroute using tunnels. IETF Internet draft, Nov 2007. http://tools.ietf.org/html/draft-bryant-ipfrr-tunnels-03.

[14] S. Bryant, M. Shand, and S. Previdi. IP fast reroute using not-via addresses. IETF Internet draft, Jul 2009. http://tools.ietf.org/html/draft-ietf-rtgwg-ipfrr-notvia-addresses-04.

[15] S. Nelakuditi, S. Lee, Y. Yu, Z.-L. Zhang, and C.-N. Chuah. Fast local rerouting for handling transient link failures. *IEEE/ACM Transactions on Networking*, 15(2):359–372, Apr 2007.

[16] Y. Zhang. The Abilene topology and traffic matrices. Online, Dec 2004. http://www.cs.utexas.edu/~yzhang/research/AbileneTM/.

[17] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson. Measuring ISP topologies with Rocketfuel. *IEEE/ACM Transactions on Networking*, 12(1):2–16, Feb 2004.

[18] W. Liu, H. T. Karaoglu, A. Gupta, M. Yuksel, and K. Kar. Edge-to-edge bailout forward contracts for single-domain Internet services. In *Proceedings of 16th IEEE International Workshop on Quality of Service*, pages 259–268, Enschede, The Netherlands, Jun 2008.

[19] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot. Traffic matrix estimation: Existing techniques and new directions. In *Proceedings of the ACM Special Interest Group on Data Communications Conference*, pages 161–174, Pittsburgh, PA, Aug 2002.