Positioning Relay Nodes in ISP Network

Meeyoung Cha†, Sue Moon†, Chong-Dae Park‡ Aman Shaikh
Dept. of Computer Science, KAIST *AT&T Labs – Research*

Abstract— **To increase reliability and robustness of mission-critical services in the event of network failures, it is often desirable and beneficial to take advantage of *path diversity* provided by the network topology. One way of achieving this inside a single Autonomous System (AS) is to use two paths between every Origin-Destination (OD) pair. One path is the default path defined by the intra-domain routing protocol running within the AS; the other path is defined as an overlay path that passes through a strategically placed *relay node* inside the AS. The key question then is how to place such relay nodes inside an AS, which is the focus of this work.**

We propose a simple greedy algorithm to find the number and positions of relay nodes such that every OD pair has an overlay path going through a relay node that is disjoint from the default path. When it is not possible to find completely disjoint overlay paths, we allow overlay paths to have overlapped links with default paths. Since overlapped links diminish the robustness of overlay paths against a single point of failure, we introduce the notion of penalty for partially disjoint paths. We apply our algorithm on an operational tier-1 ISP network and demonstrate that our method increases network-wide resiliency against a single link failure. Based on real failure scenarios obtained from the ISP network and hypothetical traffic matrix, we demonstrate that the relay nodes selected by our algorithm provide complete protection against 75.3% of failure events and allow less than 1% of traffic to be affected for 92.8% of failure events.

I. Introduction

TRAFFIC in the Internet is largely affected by link and router failures. Studies show that the impact of single or multiple network-wide changes (*e.g.*, router and link failures, fiber cuts, and link weight changes) can propagate throughout the Internet slowly (up to several seconds or minutes) [1], [2]. During this period of routing instability, large amount of traffic can be shifted from one link to another and many applications can suffer from delay, jitter, and packet loss. Such network condition changes occur rather frequently [3], [4], and persistent end-to-end connections are very likely to experience the negative impact [5]. To increase reliability and robustness of mission-critical services in the face of temporary end-to-end path outages, it is often desirable and beneficial to take advantage of *path diversity* provided by the network topology.

One way of achieving this inside a single Autonomous System (AS) is to use multiple paths between every Origin-Destination (OD) pair. One path is the default one determined by the intra-domain routing protocol running within the AS, such as OSPF [6] or IS-IS [7]. The other path is defined as an overlay path that passes through a strategically placed *relay node* inside the AS. The key question then is how to place such relay nodes inside an AS, which is the focus of this work.

Previous work on overlay routing has focused on selecting good relay nodes based on measured metrics or QoS constraints [8], [9], [10], assuming relay nodes are already deployed. Here, we take a different viewpoint: as an ISP (Internet Service Provider), we consider the problem of *positioning relay nodes well*. An ISP can set up relay nodes inside their network and offer relaying packets as a value-added service. The focus of this work is to find a *small* set of relay nodes that offer *as much path diversity as possible* to all OD pairs.

In this work, we propose a simple greedy algorithm that finds disjoint overlay paths and evaluate our algorithm using topology of an operational tier-1 ISP network. Since a single data set has its own peculiarities and is not general enough, we use synthetic and inferred topologies in our complete evaluation. However, we limit ourselves to a tier-1 ISP network for this poster. For realistic deployment analysis, we also use a six-month period event logs from the ISP network and evaluate our method.

II. Relay Node Positioning Problem

We take a graph-theoretic approach in viewing a network. A network can be modeled as a graph $G(V, E)$, where V is a set of nodes and E is a set of directed links between pairs of nodes. A path is a finite non-null sequence of nodes and links between a pair of nodes. We term the start node of a path as an origin, the end node as a destination, and the node pair as an OD (Origin-Destination) pair. Every link in the network is assigned a weight, and the cost of a path is measured as the sum of the weights of all links along the path. We limit our study to intra-domain routing and assume that Shortest Path First (SPF) routing in terms of link weights is used between an OD pair. If two paths do not have any common link between them, we call them *disjoint*.

We define the relay node positioning problem as follows. Given a network, we want to determine the number and positions of relay nodes such that every OD pair has two paths between the origin and the destination: one path is the normal routing path (termed as the "default path"), and the other path goes via one of the relay nodes (termed as the "overlay path"). When two paths are used, network is considered *resilient* as long as either one of the paths is not affected by a failure. By using two disjoint paths, we aim at providing enhanced robustness against a single link failure in the network.

A. Practice of ECMPs in a Tier-1 ISP Network

Before we introduce our method to find relay nodes for disjoint overlay paths, we introduce how path diversity is charac-

†These authors are supported by Korea Science and Engineering Foundation (KOSEF) through Advanced Information Technology Research Center (AITrc). ‡This author is supported by Brain Korea (BK) 21 Project and the school of information technology in KAIST.

terized in a typical tier-1 ISP network. Studies [11], [12] show that a typical tier-1 ISP network has a significant level of path diversity in its IP layer topology. Figure 1 shows an example of a typical tier-1 ISP network. A large ISP network consists of a collection of physical locations called Point-of-Presences, or PoPs. Within a PoP, an access router (denoted as AR_n) is connected to two or more backbone routers (denoted as BR_n) with equal link weights for fault tolerance. Typically, parallel links between a pair of two PoPs are assigned the same weight. As a result, multiple shortest paths exist between access routers in two PoPs, and they are called *Equal Cost Multi-Paths (ECMPs)*.

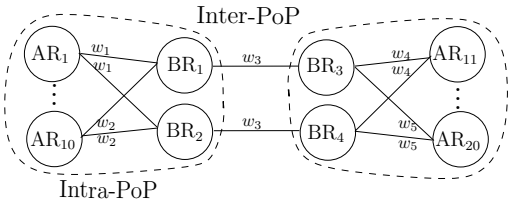

Fig. 1. Path diversity in tier-1 ISP network.

B. Impact of ECMPs on Overlay Path Selection

Under ECMP, a naive idea to provide disjoint overlay path is to avoid using all such paths. Since each node has finite degree, a node pair with ECMP may fail to have completely disjoint overlay paths. For such OD pair, we are forced to have overlapped links between default path and overlay path. Overlapped links will diminish robustness since a network is less resilient to link or router failures. For this, we introduce a measure of penalty based on the number of overlapped links and the fraction of traffic carried on those links in the next section.

III. Computing Relay Nodes for Disjoint Paths

In this section, we introduce the key concepts of our algorithm: the number of overlapped links in overlay path and a measure of penalty for limited resiliency against a single link failure. First of all, we consider a way to quantify the impact of a particular link failure. We use notation $o \rightarrow d$ to denote a shortest path from node o to d. We define an indicator variable $\mathcal{I}_{o,d,l}$ as the conditional probability that the path $o \rightarrow d$ fails given that link l fails.

$$\mathcal{I}_{o,d,l} = P[o \rightarrow d \text{ fails} \mid l \text{ fails}] \tag{1}$$

The indicator variable reflects the impact of a particular link failure on a given path. When the value is 1, it means that path $o \rightarrow d$ will certainly fail if link l fails. In other words, l is always included in $o \rightarrow d$. Otherwise, if l is not used on any path of $o \rightarrow d$, $\mathcal{I}_{o,d,l}$ is 0. In this case, failure of l does not have any impact on $o \rightarrow d$. When the value is $0 < p < 1$, it means that some path of $o \rightarrow d$ includes l and others do not. Therefore, $o \rightarrow d$ will fail with probability p if l fails. This happens when ECMPs exist. We say $o \rightarrow d$ is *affected* by a link failure on l if $\mathcal{I}_{o,d,l} > 0$.

Since the link failure of l affects $o \rightarrow d$ with probability $\mathcal{I}_{o,d,l}$, the expected number of links that affect $o \rightarrow d$ is calculated by (2). When there is no ECMP in a network, variable (2) can be interpreted as the number of links in $o \rightarrow d$. When there are multiple shortest paths, variable (2) can be interpreted as the expected number of links in $o \rightarrow d$.

$$\sum_{\forall l} \mathcal{I}_{o,d,l} \tag{2}$$

Now we consider the case when overlay path is used along with default path. Assume that a link l is used in both default path, $o \rightarrow d$, and overlay path, $o \rightarrow r \rightarrow d$. Then, the failure of link l affects both the paths. If l is used in only one of the paths, the failure of l does not affect the other path. That is, a path between o and d is resilient to the failure of l. Therefore, we consider the number of overlapped links between default path and overlay path as a measure of penalty, and it is opposite of network resiliency against a single link failure. The expected number of common links between $o \rightarrow d$ and $o \rightarrow r \rightarrow d$ is calculated,

$$\sum_{\forall l} \mathcal{I}_{o,d,l}[1 - (1 - \mathcal{I}_{o,r,l})(1 - \mathcal{I}_{r,d,l})]. \tag{3}$$

If variable (3) is 0, it implies that the overlay path is completely disjoint from the default path. For network resiliency against a single link failure, we propose selecting relay nodes such that variable (3) is minimized for all OD pairs. We formulate the positioning of relay nodes as the following. We assume that every node can be used as a relay node. Given a graph of nodes and links with weights, we first find all possible candidates of relay nodes for each OD pair such that variable (3) is within a threshold value [1]. Then out of all the relay node candidates, we choose a *small* subset that provide overlay paths for all OD pairs. Choosing a subset of the relay node candidates is a classic set covering problem which is known to be NP-hard, and we use a simple greedy approximation to find a small set of relay nodes [13]. In the next section, we give preliminary results on how relay nodes selected by our algorithm increase network resiliency in a real network topology.

IV. Evaluation of Algorithm in a Tier-1 ISP

To evaluate the algorithm, we use the topology and event logs of an **operational tier-1 ISP backbone**. The topology used has about 100 routers and 200 links. About 39% of all OD pairs have ECMPs and about 53% pairs fail to have completely disjoint overlay paths [2]. The event log used in the evaluation spans a six-month period from June 1 to November 30, 2004. The log contains five types of events: router up, router down, link up, link down, and link weight changes. When a router comes up or goes down, all links incident on the router also come up or down. Link or router down events usually cause traffic loss for some OD pairs, resulting in service disruption. On the other hand, for router/link up and weight change events, shortest paths are recomputed and some OD pairs may experience a route change (or a traffic shift) in their default paths, however detrimental impact of such a change is smaller compared to link or router failures [2]. Therefore, we only focus on link and router failure events in this section. It should be noted though that our algorithm is applicable and effective against routing instability caused by link or router up events as well.

[1] Throughout the simulation in this work, we use a threshold value of 3 allowing at most 3 overlapped links.

[2] Such failures may be due to ECMP, pathological link weight assignment, or topological characteristics.

We assume that each node re-calculates its routes immediately and instantaneously after each event. We realize this assumption by updating the topology and recomputing the shortest paths after each event. Relay nodes, when used in the analysis, are chosen based on the topology snapshot at the beginning of the event log (*i.e.*, June 1st 2004), and are kept unchanged even though the topology changes as events unfold.

To determine the impact of failure events on traffic, we assume that equal amount of traffic flows between origin and destination of every OD pair. For each event (single/multiple link and/or router failures), we calculate the fraction of this hypothetical traffic lost due to the failure with and without relay nodes. When the shortest path between an OD pair (o, d) contains the failed link l, we assume that the fraction of traffic assigned to that particular link, $\mathcal{I}_{o,d,l}$, is *lost*. In this way, we determine the fraction of traffic lost due to the failure for every OD pair.

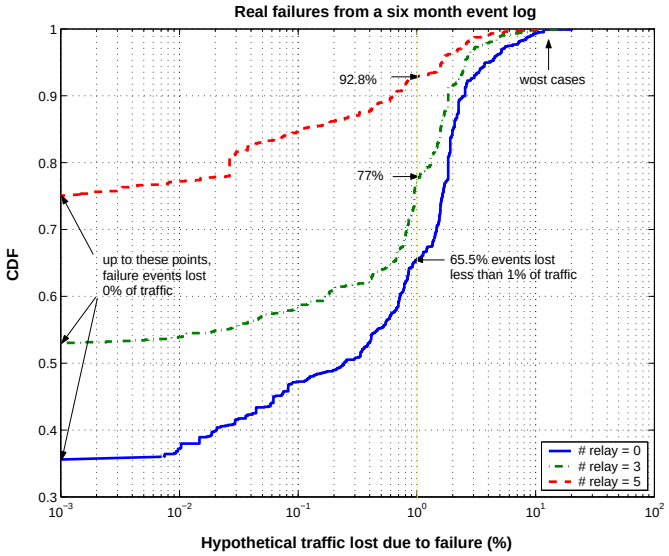

Fig. 2. Impact of failure events on traffic with and without relay nodes for a tier-1 ISP network.

Figure 2 shows a CDF log-plot of hypothetical traffic lost for the event log. The plot shows three graphs. The first one (denoted by a solid line) shows traffic loss when only default paths are used (*i.e.*, # relay = 0). The second graph (denoted by a dash-dotted line) uses both default and overlay paths with three relay nodes (*i.e.*, # relay = 3). Finally, the third graph (denoted by a dashed line) shows traffic loss with five relay nodes (*i.e.*, # relay = 5). When only default paths are used, 35.9% of failure events have no impact on the hypothetical traffic. Detailed analysis of these events show that link weights are set to a large value before the corresponding link goes down. Setting a link weight to a high value forces the traffic to bypass the link, allowing a "graceful" link shutdown. The remaining events impact only a small fraction of traffic in the network; for 65.5% of failure events, less than 1% of hypothetical traffic is lost.

When three relay nodes are used, they provide complete resilience against 52.9% of failure events, which is a 17% increase when compared against no relay node case. Better still, up to 77% of failure events affect 1% or less of hypothetical traffic. When five relay nodes are used, network resilience to real failures increases further. In this case, using overlay paths provide complete protection against 75.3% of failure events. Furthermore, for 92.8% of failure events, less than 1% of hypothetical traffic is affected. It is also worth noting that a small number of relay nodes chosen at the beginning of the period is effective in providing resilience against failures over the entire course of six months.

V. Conclusions

In this work, we propose a simple greedy algorithm for selecting the number and positions of relay nodes in a network run by a single AS. While completely disjoint overlay paths are most desirable, unfortunately, in reality, it is often not possible to find completely disjoint paths for all node pairs. In such scenarios, it is still beneficial to have overlay paths that partially overlap with the default paths. Towards that end, we propose the intuitive notion of penalty for partially disjoint overlay paths, and find relay nodes that incur least amount of penalty.

We evaluate the efficacy of our algorithm with an operational tier-1 ISP network. Based on real failure scenarios from the ISP network and hypothetical traffic matrix, we demonstrate that the relay nodes selected by our algorithm provide complete protection against 75.3% of failure events and allow less than 1% traffic to be affected for 92.8% of failure events.

References

[1] Craig Labovitz, Abha Ahuja, Abhijit Bose, and Farnam Jahanian, "Delayed Internet routing convergence," in *Proceedings of ACM SIGCOMM*, September 2000, pp. 175–187.

[2] Catherine Boutremans, Gianluca Iannaccone, and Christophe Diot, "Impact of link failures on VoIP performance," in *Proceedings of Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, May 2002.

[3] Athina Markopoulou, Gianluca Iannaccone, Supratik Bhattacharyya, Chen-Nee Chuah, and Christophe Diot, "Characterization of failures in an IP backbone," in *Proceedings of IEEE INFOCOM*, March 2004.

[4] Vern Paxson, "End-to-end routing behavior in the Internet," *IEEE/ACM Transactions on Networking*, vol. 5, no. 5, pp. 601–615, 1997.

[5] Sanghwan Lee, Yinzhe Yu, Srihari Nelakuditi, Zhi-Li Zhang, and Chen-Nee Chuah, "Proactive vs reactive approaches to failure resilient routing," in *Proceedings of IEEE INFOCOM*, 2004.

[6] J. Moy, "RFC 2328: OSPF version 2," 1998.

[7] R. W. Callon, "RFC 1195: Use of OSI IS-IS for routing in TCP/IP and dual environments," December 1990.

[8] David Andersen, Hari Balakrishnan, M. Frans Kaashoek, and Robert Morris, "Resilient overlay networks," in *Proceedings of ACM Symposium on Operating Systems Principles (SOSP)*, 2001.

[9] T. Nguyen and A. Zakhor, "Path Diversity with Forward error correction (PDF) system for packet switched networks," in *Proceedings of IEEE INFOCOM*, March 2003.

[10] Lakshminarayanan Subramanian, I. Stoica, H. Balakrishnan, and R. Katz, "OverQoS: Offering Internet QoS using overlays," in *Proceedings of HotNets–I*, October 2002.

[11] Renata Teixeira, Keith Marzullo, Stefan Savage, and Geoffrey M. Voelker, "In Search of Path Diversity in ISP Network," in *Proceedings of ACM SIGCOMM IMC*, October 2003.

[12] G. Iannaccone, C. N. Chuah, S. Bhattacharyya, and C. Diot, "Feasibility of IP restoration in a tier-1 backbone," *IEEE Network Magazine*, vol. 18, no. 2, pp. 13–19, 2004.

[13] Petr Slavík, "A tight analysis of the greedy algorithm for set cover," *Journal of Algorithms*, vol. 25, no. 2, pp. 237–254, 1997.