Cross-Layer Approach for Minimizing Routing Disruption in IP Networks

Qiang Zheng, *Student Member, IEEE*, Guohong Cao, *Fellow, IEEE*, Thomas F. La Porta, *Fellow, IEEE*, and Ananthram Swami, *Fellow, IEEE*

Abstract—Backup paths are widely used in IP networks to protect IP links from failures. However, existing solutions such as the commonly used independent model and Shared Risk Link Group (SRLG) model do not accurately reflect the correlation between IP link failures, and thus may not choose reliable backup paths. We propose a cross-layer approach for minimizing routing disruption caused by IP link failures. We develop a probabilistically correlated failure (PCF) model to quantify the impact of IP link failure on the reliability of backup paths. With the PCF model, we propose an algorithm to choose multiple reliable backup paths to protect each IP link. When an IP link fails, its traffic is split onto multiple backup paths to ensure that the rerouted traffic load on each IP link does not exceed the usable bandwidth. We evaluate our approach using real ISP networks with both optical and IP layer topologies. Experimental results show that two backup paths are adequate for protecting a logical link. Compared with existing works, the backup paths selected by our approach are at least 18 percent more reliable and the routing disruption is reduced by at least 22 percent. Unlike prior works, the proposed approach prevents the rerouted traffic from interfering with normal traffic.

Index Terms—Routing, failures, cross-layer, recovery, IP networks

1 Introduction

IP link failures are fairly common in the Internet for various reasons. In high speed IP networks like the Internet backbone, disconnection of a link for several seconds can lead to millions of packets being dropped [1]. Therefore, quickly recovering from IP link failures is important for enhancing Internet reliability and availability, and has received much attention in recent years. Currently, backup path-based protection [2], [3], and [4] is widely used by Internet Service Providers (ISPs) to protect their domains. In this approach, backup paths are precomputed, configured, and stored in routers. When a link failure is detected, traffic originally traversing the link is immediately switched to the backup path of this link. Through this, the routing disruption duration is reduced to the failure detection time which is typically less than 50 ms [5].

Selecting backup paths is a critical problem in backup path-based protection. Existing approaches mainly focus on choosing reliable backup paths to reduce the routing disruption caused by IP link failures. However, they suffer from two limitations. First, the widely used failure models do not accurately reflect the correlation between IP link failures. As a result, the selected backup paths may be unreliable. Second, most prior works consider backup path selection as a connectivity problem, but ignore the traffic load and bandwidth constraint of IP links.

Current IP backbone networks are primarily built on the Wavelength Division Multiplexing (WDM) infrastructure [6]. In this layered structure, the IP layer topology (logical topology) is embedded on the optical layer topology (physical topology), and each IP link (logical link) is mapped to a lightpath in the physical topology. An IP link may consist of multiple fiber links, and a fiber link may be shared by multiple IP links. When a fiber link fails, all the logical links embedded on it fail simultaneously. Fig. 1 shows an example of the topology mapping in IP-over-WDM networks. The logical topology in Fig. 1a is embedded on the physical topology shown in Fig. 1b, in which nodes v_5, v_6, and v_7 are optical layer devices and hence do not appear in the logical topology. Logical links are mapped to lightpaths as shown in Fig. 1c. For example, $e_{1,4}$ shares fiber link $f_{1,5}$ with $e_{1,3}$ and shares fiber link $f_{4,7}$ with $e_{3,4}$.

In prior works, logical link failures were considered as independent events [7], [8], [9], and [10] or modeled as a Shared Risk Link Group (SRLG[1]) [11], [12], and [13]. However, both models have limitations. First, logical link failures are not independent because of the topology mapping. In Fig. 1, when fiber link $f_{1,5}$ fails, logical links $e_{1,2}$, $e_{1,3}$, and $e_{1,4}$ will fail together. This shows that failures of $e_{1,2}$, $e_{1,3}$, and $e_{1,4}$ are correlated rather than independent. Second, sharing fiber links does not imply that logical links in the same SRLG must fail simultaneously. For example, $e_{1,2}$, $e_{1,3}$, and $e_{1,4}$ are in the same SRLG. When $e_{1,4}$ fails, it does not mean that $e_{1,2}$ and $e_{1,3}$ must also fail. If the failure of $e_{1,4}$ is caused by fiber link $f_{4,7}$, $e_{1,2}$ and $e_{1,3}$ may be live. In

- *Q. Zheng is with the The Pennsylvania State University, Department of Computer Science and Engineering, University Park, PA 16802 USA.*
- *G. Cao is with the The Pennsylvania State University, Department of Computer Science, University Park, PA 16802 USA.*
- *T.F. La Porta is with the Pennsylvania State University, EIC, Transactions on Mobile Computing, University Park, PA 16802 USA.*
- *A. Swami is with the Army Research Laboratory, Adelphi, MA USA.*

Manuscript received 7 Nov. 2012; revised 24 Apr. 2013; accepted 28 May 2013. Date of publication 13 June 2013; date of current version 13 June 2014.
Recommended for acceptance by J. Lloret.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPDS.2013.157

[1] A SRLG is a set of IP links that share the same risk such as a fiber link failure. If an IP link fails, all the IP links within the same SRLG are considered as failed.

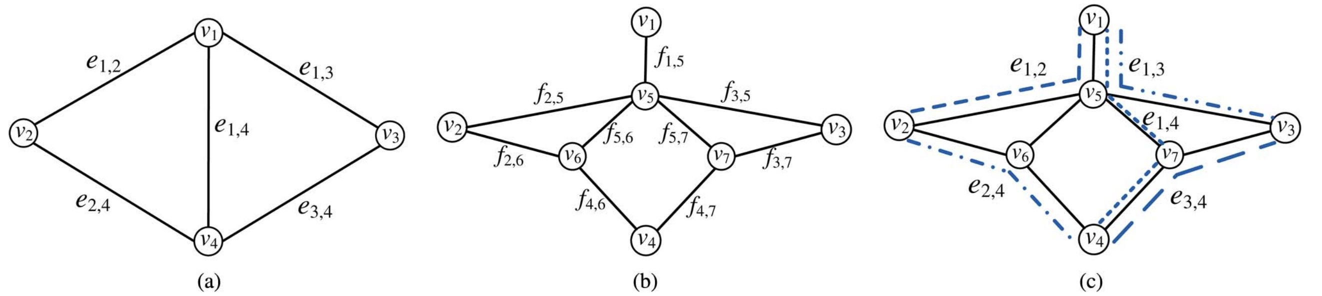


Fig. 1. Example of the mapping between the logical and physical topologies in IP-over-WDM networks. (a) Logical topology. (b) Physical topology. (c) Mapping between the logical and fiber links.

fact, recent Internet measurements [14], [15] show that independent failures and correlated failures coexist in the Internet. When $e_{1,4}$ fails, it may be an independent failure or a correlated failure due to shared fiber links. Therefore, $e_{1,2}$ and $e_{1,3}$ may also fail with a certain probability, i.e., failures of $e_{1,2}$, $e_{1,3}$, and $e_{1,4}$ are *probabilistically correlated*. This feature cannot be modeled by the traditional independent and SRLG models, and has not been investigated in backup path selection.

Most existing approaches focus on selecting reliable backup paths, but ignore the fact that a backup path may not have enough bandwidth for the rerouted traffic. As a result, the rerouted traffic load on some logical links may exceed their usable bandwidth, and thus cause link overload. As Iyer *et al.* observed [16], most link overload in an IP backbone is caused by the traffic rerouted due to IP link failures. In a survey in 2010, two of the largest ISPs in the world reported congestion caused by rerouted traffic in their networks [17]. Therefore, backup paths should be carefully selected to prevent causing link overload.

We propose a cross-layer approach to minimize routing disruption caused by IP link failures. The basic idea is to consider the correlation between IP link failures in backup path selection and protect each IP link with multiple reliable backup paths. A key observation is that the backup path for an IP link is used only when the IP link fails. Therefore, the reliability of backup path should be calculated under the condition that the IP link fails. We develop a probabilistically correlated failure (PCF) model based on the topology mapping and the failure probability of fiber links and logical links. The PCF model calculates the failure probability of fiber links, logical links, and backup paths under the condition that an IP link fails. Hence, we can determine reliable backup paths with the PCF model. With the PCF model, we propose an algorithm to select at most N reliable backup paths for each IP link and compute the rerouted traffic load on each backup path. This ensures that the rerouted traffic load on each IP link does not exceed its usable bandwidth so as to avoid link overload.

Our approach is different from prior works in three aspects. First, it is based on a cross-layer design, which considers the correlation between logical and physical topologies. The proposed PCF model can reflect the probabilistic correlation between logical link failures. Second, we protect each logical link with multiple backup paths to effectively reroute traffic and avoid link overload, whereas most prior works select single backup path for each logical link. Third, our approach considers the traffic load and bandwidth constraint. It guarantees that the rerouted traffic load does not exceed the usable bandwidth, even when multiple logical links fail simultaneously. We evaluate the proposed approach using real ISP networks with both optical and IP layer topologies. Experimental results show that two backup paths are adequate for protecting a logical link. Compared with existing works, the backup paths selected by our approach are at least 18 percent more reliable and the routing disruption is reduced by at least 22 percent. Moreover, the proposed approach prevents logical link overload caused by the rerouted traffic.

The rest of this paper is organized as follows. Section 2 presents the background for our work. Section 3 presents the PCF model. Section 4 describes the backup path selection problem and solution. Section 5 presents the performance evaluation and Section 6 reviews related work. Finally, Section 7 concludes the paper.

2 Preliminaries

This section introduces backup path-based IP link protection and a model of IP-over-WDM networks.

2.1 Backup Path-Based IP Link Protection

In the current Internet, each router monitors the connectivity with its neighboring routers. When a logical link fails, only the two routers connected by it can detect the failure. Hence, a router may not have the overall information of failures in the network. Although the failed logical links can be identified within a few seconds [18], this waiting time translates to a lot of dropped packets on a high bandwidth optical link. As a result, a recovery approach cannot wait until finishing collecting the overall information of failures and then reroute traffic.

Instead, backup paths are widely used to quickly reroute the traffic affected by failures. In backup path-based IP link protection, a router precomputes backup paths for each of its logical links. On detecting a link failure, the router immediately switches the traffic originally sent on that logical link onto the corresponding backup paths. After the routing protocol converges to a new network topology, routing paths will not contain the failed logical link and the router has a reachable next hop for each destination. Therefore, the router stops using the backup path to reroute traffic. Moreover, routers recompute backup paths based on the new network topology. Backup paths can be implemented with Multi-Protocol Label Switching [19] which is widely supported in the current Internet. Each

TABLE 1
Table of Notations

Symbols	Meaning
$a_{m,n}^{i,j}$	1 if $e_{i,j}$ is embedded on $f_{m,n}$, 0 otherwise
$B_{i,j}^{k}$	the kth backup path of $e_{i,j}$
$c_{i,j}$	capacity of $e_{i,j}$
$D_{i,j}$	traffic disruption of $e_{i,j}$
D	routing disruption of the whole network
$e_{i,j}$	the logical link from $v_i \in V_L$ to $v_j \in V_L$
$f_{m,n}$	the fiber link from $v_m \in V_P$ to $v_n \in V_P$
$F_{i,j}$	fiber links shared by $e_{i,j}$ and other logical links
$G_L = (V_L, E_L)$	logical topology
$G_P = (V_P, F_P)$	physical topology
$l_{i,j}$	traffic load on $e_{i,j}$ under normal conditions
N	each logical link has at most N backup paths
$p_{i,j}$	unconditional failure probability of $e_{i,j}$
$p_{i,j}^{C}$	correlated failure probability of $e_{i,j}$
$p_{i,j}^{I}$	independent failure probability of $e_{i,j}$
$P(B_{i,j}^{k} \vert e_{i,j})$	failure probability of $B_{i,j}^{k}$ when $e_{i,j}$ fails
$P(e_{s,t} \vert e_{i,j})$	failure probability of $e_{s,t}$ when $e_{i,j}$ fails
$P(f_{m,n} \vert e_{i,j})$	failure probability of $f_{m,n}$ when $e_{i,j}$ fails
$q_{m,n}$	unconditional failure probability of $f_{m,n}$
$r_{i,j}^{k}$	rerouted traffic load on $B_{i,j}^{k}$
$x_{s,t}^{i,j,k}$	1 if $B_{i,j}^{k}$ uses $e_{s,t}$, 0 otherwise

backup path is configured as a Label-Switched Path (LSP) and the rerouted traffic can be split on backup paths.

2.2 Model of IP-over-WDM Networks

Backup path-based protection is primarily used for introdomain routing, which is really deployed by ISPs to protect their domains. Similarly, our approach is also for intra-domain routing. ISPs instrument their networks heavily and have the network topology, link capacity, and traffic demands which are used in our approach. The IP-over-WDM network under study has a logical topology and a physical topology, which are commonly modeled as two undirected graphs [20], [21], [22], and [23]. In the physical topology $G_P = (V_P, F_P)$, V_P is a set of nodes and F_P is a set of fiber links. The fiber link from $v_m \in V_P$ to $v_n \in V_P$ is denoted by $f_{m,n}$. In the logical topology $G_L = (V_L, E_L)$, $V_L \subseteq V_P$ and E_L is a set of logical links. The logical link from $v_i \in V_L$ to $v_j \in V_L$ is denoted by $e_{i,j}$. Each logical link is mapped on the physical topology as a lightpath, i.e., a path over the fiber links. Hence a logical link is *embedded* on fiber links, or a fiber link *carries* logical links. The topology mapping is established during network configuration, and thus is known to us. Unlike logical link states, the topology mapping is quite stable and does not frequently change. When the network administrator adjusts the topology mapping, the topology mapping information at routers can also be updated. Table 1 summarizes the notations used in this paper.

3 Probabilistically Correlated Failure Model

This section describes the probabilistically correlated failure (PCF) model. We first provide some motivation and then present the details of this model. We also provide an example in Appendix A, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPDS.2013.157, to show how the model works and differentiate it from the traditional independent and SRLG models.

3.1 Motivation

Recent measurements [14], [15] show that there are two types of IP link failures in the Internet, i.e., *independent failures* and *correlated failures*. Independent failures are unrelated. They occur for several reasons, such as hardware failures, configuration errors, and software bugs. Correlated failures are mainly caused by failures of fiber links carrying multiple logical links. When a logical link has a correlated failure, it implies that some other logical links sharing fiber links with it may also fail.

Since each router only monitors the connectivity with its neighboring routers, routers cannot determine whether a logical link failure is independent or correlated. The failure of $e_{i,j}$ implies that the logical links sharing at least one fiber link with $e_{i,j}$ may also fail with a certain probability. Therefore, backup path selection approaches should consider this probabilistic correlation between logical link failures. However, the traditional independent and SRLG models take the correlation between logical link failures as a *none-or-all* relation. The independent model considers that logical links only have independent failures and thus it usually underestimates the failure probability of logical links; whereas the SRLG model considers that logical links only have correlated failures and usually overestimates the failure probability.

We develop a PCF model based on the topology mapping and the failure probability of fiber links and logical links. The PCF model considers the probabilistic relation between logical link failures. The objective is to quantify the impact of a logical link failure on the failure probability of other logical links and backup paths. With the PCF model, we propose an algorithm to choose reliable backup paths to minimize the routing disruption.

3.2 The PCF Model

The PCF model is built on three kinds of information, i.e., the topology mapping, failure probability of fiber links, and failure probability of logical links, all of which are already gathered by ISPs. ISPs configure their topology mapping, and thus they have this information. The failure probability of fiber links and logical links can be obtained with Internet measurement approaches [14], [15] deployed at the optical and IP layers. Monitoring mechanisms at the optical layer can detect fiber link failures through SONET alarms. The information of logical link failures can be extracted from routing updates. ISPs also maintain failure information, because they monitor the optical and IP layers of their networks.

A key observation is that the failure probability of the backup paths for logical link $e_{i,j}$ should be computed under the condition that $e_{i,j}$ fails, because the backup paths are used only when $e_{i,j}$ fails. A backup path is built on logical links, and a logical link is embedded on fiber links. Hence, we first compute the failure probability of fiber links under the condition that $e_{i,j}$ fails. Then, we compute the conditional failure probability of logical links and backup paths.

The unconditional failure probability of logical link $e_{i,j}$ is denoted by $p_{i,j} \in [0, 1)$, which includes independent and correlated failures. However, it cannot reveal the correlation between logical link failures and thus we cannot directly use it to compute the conditional failure probability of backup paths. Unlike logical links, most fiber link failures are independent [13], [23]. We assume that a fiber link $f_{m,n}$ fails independently with probability $q_{m,n} \in [0, 1)$. In practice, we may obtain $p_{i,j}$ and $q_{m,n}$ based on previous logical link failures and fiber link failures. Let $a^{i,j}_{m,n}$ defined in Eq. (1) express the mapping between logical link $e_{i,j}$ and fiber link $f_{m,n}$

$$a^{i,j}_{m,n} = \begin{cases} 1 & \textbf{if } e_{i,j} \textbf{ is embedded on } f_{m,n} \\ 0 & \textbf{otherwise.} \end{cases} \tag{1}$$

A logical link is subject to failures and correlated failures are caused by the fiber links carrying multiple logical links. Let $F_{i,j}$ be the set of fiber links shared by $e_{i,j}$ and other logical links. $F_{i,j}$ is defined by Eq. (2). Suppose that a fiber link $f_{m,n}$ carries $e_{i,j}$, i.e., $a^{i,j}_{m,n} = 1$. If there is another logical link $e_{s,t}$ that is also carried by $f_{m,n}$, $f_{m,n}$ is in the set $F_{i,j}$

$$F_{i,j} = \Big\{ f_{m,n} | a^{i,j}_{m,n} a^{s,t}_{m,n} = 1, e_{i,j} \in E_L, \exists e_{s,t} \in E_L, \\ \forall f_{m,n} \in F_P \Big\}. \tag{2}$$

Let $p^C_{i,j}$ be the probability that $e_{i,j}$ has correlated failures with other logical links. Since fiber link failures are independent, $p^C_{i,j}$ is computed by Eq. (3). If $e_{i,j}$ does not share a fiber link with other logical links, its correlated failure probability is 0

$$p^C_{i,j} = \begin{cases} 0 & \textbf{if } F_{i,j} = \emptyset \\ 1 - \prod_{f_{m,n} \in F_{i,j}} (1 - q_{m,n}) & \textbf{otherwise.} \end{cases} \tag{3}$$

Suppose the independent failure probability of $e_{i,j}$ is $p^I_{i,j}$. We have the relation shown in Eq. (4), because failures of $e_{i,j}$ are either independent or correlated. Lemma 1 shows that $p^I_{i,j}$ must be strictly less than unity

$$p_{i,j} = 1 - \left(1 - p^I_{i,j}\right)\left(1 - p^C_{i,j}\right). \tag{4}$$

Lemma 1. $p^I_{i,j} \in [0, 1)$.

Proof. According to Eq. (4), $p^I_{i,j} = \frac{p_{i,j} - p^C_{i,j}}{1 - p^C_{i,j}}$. Since $q_{m,n} \in [0, 1)$ for each fiber link $q_{m,n}$, we have $p^C_{i,j} \in [0, 1)$. Together with $p_{i,j} \in [0, 1)$, we have $p^I_{i,j} < 1$. Therefore, $p^I_{i,j} \in [0, 1)$. □

Based on the above information, we compute the failure probability of fiber links under the condition that $e_{i,j}$ fails. Let $P(f_{m,n}|e_{i,j})$ be this conditional failure probability. We only need to deal with the fiber link $f_{m,n}$ with failure probability $q_{m,n} > 0$. There are three cases as follows.

- Case 1: $e_{i,j}$ is not embedded on $f_{m,n}$. It means that the failure of $e_{i,j}$ is not related with $f_{m,n}$. Hence, $P(f_{m,n}|e_{i,j})$ is equal to $q_{m,n}$.
- Case 2: $f_{m,n}$ only carries $e_{i,j}$, then a failure of $f_{m,n}$ leads to an independent failure of $e_{i,j}$. In this case, $P(f_{m,n}|e_{i,j})$ is calculated by Eq. (5), where $P(e^I_{i,j}|e_{i,j})$ is the probability that $e_{i,j}$ has an independent failure when it fails, and $P(f_{m,n}|e^I_{i,j})$ is the probability that this independent failure is caused by $f_{m,n}$

$$P(f_{m,n}|e_{i,j}) = P\left(e^I_{i,j}|e_{i,j}\right) \times P\left(f_{m,n}|e^I_{i,j}\right). \tag{5}$$

Based on Bayes' theorem, $P(e^I_{i,j}|e_{i,j})$ is calculated by Eq. (6). $P(e_{i,j}|e^I_{i,j})$ is the failure probability of $e_{i,j}$ when its independent failures occur, which is 1. $P(e^I_{i,j})$ is the independent failure probability of $e_{i,j}$, i.e., $p^I_{i,j}$. $P(e_{i,j})$ is the failure probability of $e_{i,j}$, i.e., $p_{i,j}$

$$\begin{aligned} P\left(e^I_{i,j}|e_{i,j}\right) &= \frac{P\left(e_{i,j}|e^I_{i,j}\right) P\left(e^I_{i,j}\right)}{P(e_{i,j})} \\ &= \frac{P\left(e^I_{i,j}\right)}{P(e_{i,j})} = \frac{p^I_{i,j}}{p_{i,j}}. \end{aligned} \tag{6}$$

Similarly, $P(f_{m,n}|e^I_{i,j})$ is computed by Eq. (7). Since $q_{m,n}$ is not 0, $p^I_{i,j}$ cannot be 0

$$P\left(f_{m,n}|e^I_{i,j}\right) = \frac{P\left(e^I_{i,j}|f_{m,n}\right) P(f_{m,n})}{P\left(e^I_{i,j}\right)} = \frac{q_{m,n}}{p^I_{i,j}}. \tag{7}$$

Based on Eqs. (5), (6), (7), $P(f_{m,n}|e_{i,j}) = \frac{q_{m,n}}{p_{i,j}}$.

- Case 3: $f_{m,n}$ carries $e_{i,j}$ and some other logical links, then a failure of $f_{m,n}$ leads to correlated failures. The calculation of $P(f_{m,n}|e_{i,j})$ is similar to that in case 2, in which $e^I_{i,j}$ is replaced by $e^C_{i,j}$ and $p^I_{i,j}$ is replaced by $p^C_{i,j}$. The result is the same as in case 2, i,e., $P(f_{m,n}|e_{i,j}) = \frac{q_{m,n}}{p_{i,j}}$.

In summary, the conditional failure probability $P(f_{m,n}|e_{i,j})$ is given by Eq. (8). It is only defined for $e_{i,j}$ whose failure probability $p_{i,j}$ is not 0. If $p_{i,j}$ is 0, $e_{i,j}$ never fails, and thus we do not need to select backup paths for it

$$P(f_{m,n}|e_{i,j}) = \begin{cases} q_{m,n} & a^{i,j}_{m,n} = 0 \\ \frac{q_{m,n}}{p_{i,j}} & a^{i,j}_{m,n} = 1. \end{cases} \tag{8}$$

The conditional probability cannot be smaller than the unconditional probability. The following Lemma 2 shows that conditioning strictly increases the probability of fiber links.

Lemma 2. *If $f_{m,n}$ carries $e_{i,j}$ and $q_{m,n} > 0$, $P(f_{m,n}|e_{i,j}) > q_{m,n}$.*

Proof. If $f_{m,n}$ carries $e_{i,j}$, the second case of Eq. (8) holds, i.e., $P(f_{m,n}|e_{i,j}) = \frac{q_{m,n}}{p_{i,j}}$. Since $q_{m,n} > 0$, the failure probability of $e_{i,j}$ is above 0. Since $p_{i,j} \in [0, 1)$, we have $p_{i,j} \in (0, 1)$. Therefore, $P(f_{m,n}|e_{i,j}) = \frac{q_{m,n}}{p_{i,j}} > q_{m,n}$. □

Lemma 2 indicates the condition under which the posterior probability of $f_{m,n}$ is increased due to a failure of $e_{i,j}$. Accordingly, the failure probability of the logical links embedded on $f_{m,n}$ is affected by the failure of $e_{i,j}$.

Next, we calculate the failure probability of logical links under the condition that $e_{i,j}$ fails. A logical link $e_{s,t}$ has independent failures and correlated failures. Its independent failure probability is $p^I_{s,t}$, whether or not $e_{i,j}$ fails. However, its correlated failure probability may be affected by the failure of $e_{i,j}$. We use $P(e^C_{s,t}|e_{i,j})$ to denote the correlated failure probability of $e_{s,t}$ under the condition that $e_{i,j}$ fails. It is calculated by Eq. (9) (see also Eq. (3)). Set $F_{s,t}$ contains all fiber links shared by $e_{s,t}$ and other logical links. If $e_{s,t}$ does not share a fiber link with other logical links (i.e., $F_{s,t}$ is empty), $e_{s,t}$ does not have correlated failures and thus its correlated failure probability under condition of $e_{i,j}$ failure is 0. If $F_{s,t}$ is not empty, $P(e^C_{s,t}|e_{i,j})$ is computed using

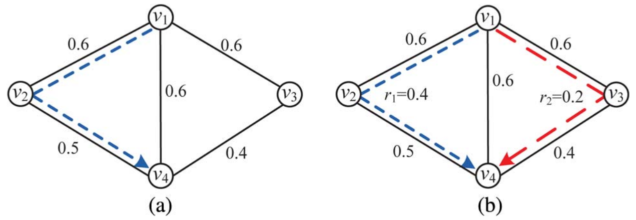


Fig. 2. Motivation for protecting a logical link with multiple backup paths. (a) Single backup path may not have enough bandwidth. (b) The rerouted traffic is split on two backup paths.

each fiber link $f_{m,n}$ in $F_{s,t}$, because only these fiber links are related with the correlated failure of $e_{s,t}$

$$P\left(e^C_{s,t}|e_{i,j}\right) = \begin{cases} 0 & \text{if } F_{s,t} = \emptyset \\ 1 - \prod_{f_{m,n}\in F_{s,t}} (1 - P(f_{m,n}|e_{i,j})) & \text{otherwise.} \end{cases} \tag{9}$$

Based on this, Eq. (10) computes the failure probability of $e_{s,t}$ under the condition that $e_{i,j}$ fails, which is denoted by $P(e_{s,t}|e_{i,j})$ (see also Eq. (4))

$$P(e_{s,t}|e_{i,j}) = 1 - \left(1 - p^I_{s,t}\right)\left(1 - P\left(e^C_{s,t}|e_{i,j}\right)\right). \tag{10}$$

Theorem 1 shows that conditioning strictly increases the probability of logical links.

Theorem 1. *$P(e_{s,t}|e_{i,j}) > p_{s,t}$, if there is a fiber link $f_{m,n}$ satisfying $a^{s,t}_{m,n} = 1$, $a^{i,j}_{m,n} = 1$, and $q_{m,n} > 0$.*

Proof. Since $p^I_{s,t} \in [0, 1)$ as shown by Lemma 1, $1 - p^I_{s,t} \in (0, 1]$. According to Eq. (4) and Eq. (10), we need to prove $P(e^C_{s,t}|e_{i,j}) > p^C_{s,t}$ to show $P(e_{s,t}|e_{i,j}) > p_{s,t}$. With Eq. (3) and Eq. (9), it is equivalent to proving the following relation:

$$\prod_{f_{m,n}\in F_{i,j}} (1 - P(f_{m,n}|e_{i,j})) < \prod_{f_{m,n}\in F_{i,j}} (1 - q_{m,n}).$$

If $q_{m,n}$ is 0, $P(f_{m,n}|e_{i,j})$ is also 0, and thus $1 - P(f_{m,n}|e_{i,j}) = 1$ and $1 - q_{m,n} = 1$. If $f_{m,n}$ carries $e_{i,j}$ and $q_{m,n} > 0$, Lemma 2 shows that $P(f_{m,n}|e_{i,j}) > q_{m,n}$. Therefore, $\prod_{f_{m,n}\in F_{i,j}} (1 - P(f_{m,n}|e_{i,j})) < \prod_{f_{m,n}\in F_{i,j}} (1 - q_{m,n})$, if there is a fiber link $f_{m,n}$ shared by $e_{i,j}$ and $e_{s,t}$ and $q_{m,n} > 0$. It means that $P(e_{s,t}|e_{i,j}) > p_{s,t}$.

As shown by Theorem 1, if logical link $e_{s,t}$ shares a fiber link $f_{m,n}$ with $e_{i,j}$ and $q_{m,n}$ is not 0, the failure probability of $e_{s,t}$ is increased due to the failure of $e_{i,j}$. Accordingly, the failure probability of the backup paths built on $e_{s,t}$ is increased when $e_{i,j}$ fails.

Next, we calculate the failure probability of backup paths under the condition that $e_{i,j}$ fails. Let $B^k_{i,j}$ denote the kth backup path of $e_{i,j}$, and $P(B^k_{i,j}|e_{i,j})$ denote the failure probability of $B^k_{i,j}$ when $e_{i,j}$ fails. The variable $x^{i,j,k}_{s,t} \in \{0, 1\}$ is defined as follows to express if $B^k_{i,j}$ uses the logical link $e_{s,t}$:

$$x^{i,j,k}_{s,t} = \begin{cases} 1 & \text{if } B^k_{i,j} \text{ uses } e_{s,t} \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

$P(B^k_{i,j}|e_{i,j})$ is computed by Eq. (12), which enumerates logical links and counts the ones traversed by $B_{i,j}$

$$P\left(B^k_{i,j}|e_{i,j}\right) = 1 - \prod_{e_{s,t}\in E_L} \left(1 - x^{i,j,k}_{s,t} P(e_{s,t}|e_{i,j})\right). \tag{12}$$

We provide an example in Appendix A which can be found in the online supplemental material to show how the PCF model works and the differences from the independent and SRLG models.

4 Backup Path Selection

With the PCF model, we propose an algorithm to select multiple backup paths to protect each IP link. Our algorithm considers both reliability and bandwidth constraints. It aims at minimizing routing disruption by choosing reliable backup paths and splitting the rerouted traffic onto them. Furthermore, it controls the rerouted traffic load to prevent causing logical link overload.

4.1 Motivation

Most existing protection approaches focus on choosing reliable backup paths, but ignore the fact that a backup path may not have enough bandwidth for the rerouted traffic. Without considering the bandwidth constraint, they commonly choose one backup path to protect each logical link. Our approach considers both reliability and bandwidth constraint. It protects each logical link with multiple backup paths and splits the rerouted traffic onto them, because there may be no individual backup path that has enough bandwidth for the rerouted traffic.

Fig. 2 illustrates the need for protecting a logical link with multiple backup paths. Logical links have capacity 1, and the number on a logical link is its traffic load under normal conditions. In Fig. 2a, v_1 uses a single backup path to protect $e_{1,4}$, whose usable bandwidth is $\min\{1 - 0.6, 1 - 0.5\} = 0.4$. When $e_{1,4}$ fails, the total traffic load on $e_{1,2}$ will exceed its bandwidth, and hence link overload occurs. Our approach protects $e_{1,4}$ with two backup paths as shown in Fig. 2b. When $e_{1,4}$ fails, the rerouted traffic load split onto the left one is 0.4, and that onto the right one is 0.2. With this approach, the entire traffic of $e_{1,4}$ can be rerouted without causing link overload.

Using more backup paths to protect a logical link is helpful for rerouting traffic, but increases the time for computing backup paths, configuration complexity, and storage overhead. We require that each logical link can have at most N backup paths. An appropriate N should be chosen based on usable network resources. In Section 5, we will investigate the impact of parameter N on the performance of our approach.

4.2 Problem Definition and Formulation

We consider both reliability of backup paths and bandwidth constraint of logical links. The objective is to minimize the routing disruption of the entire network, which was also the major objective in prior works. Furthermore, the rerouted traffic load on a logical link should not exceed its usable bandwidth to prevent logical link overload and interfering with normal traffic. This constraint is ignored by existing approaches.

First, we define routing disruption based on the PCF model. Suppose the capacity of logical link $e_{i,j}$ is $c_{i,j}$. Under normal conditions, the traffic load on $e_{i,j}$ is $l_{i,j}$ which satisfies $l_{i,j} \leq c_{i,j}$. Network administrators configure the link cost to achieve traffic engineering goals, and hence traffic load $l_{i,j}$ is known. The bandwidth of $e_{i,j}$ that can be used by backup paths is $c_{i,j} - l_{i,j}$. $e_{i,j}$ is *overloaded* if the rerouted traffic load on it exceeds $c_{i,j} - l_{i,j}$. The kth backup path is denoted by $B^k_{i,j}$, and the bandwidth reserved for it is $r^k_{i,j}$, which is the traffic load split onto $B^k_{i,j}$ when $e_{i,j}$ fails. The traffic load of $e_{i,j}$ protected by backup paths is $\sum_{k=1}^{N} r^k_{i,j}$, and the unprotected traffic load is $l_{i,j} - \sum_{k=1}^{N} r^k_{i,j}$. When $e_{i,j}$ fails, the unprotected traffic is disrupted. If $B^k_{i,j}$ also fails, the traffic rerouted on $B^k_{i,j}$ is disrupted. Therefore, the traffic disruption of $e_{i,j}$ is defined in Eq. (13), which is the mathematical expectation of the disrupted traffic load

$$D_{i,j} = p_{i,j}\left(\sum_{k=1}^{N} P\left(B^k_{i,j}|e_{i,j}\right) r^k_{i,j} + l_{i,j} - \sum_{k=1}^{N} r^k_{i,j}\right). \tag{13}$$

The routing disruption of the entire network is then defined in Eq. (14), which is the mathematical expectation of traffic disruption in the entire network

$$D = \sum_{e_{i,j} \in E_L} D_{i,j}. \tag{14}$$

The backup path selection problem is described in Definition 1.

Definition 1 (Problem Definition). *We aim to select at most N backup paths for each logical link and compute the rerouted traffic load for each backup path, such that (1) the routing disruption of the entire network is minimized; (2) the rerouted traffic load on each logical link does not exceed its usable bandwidth.*

Our problem can be formally defined as an optimization problem as shown in Eqs. (15), (16), (17), (18), (19), in which $x^{i,j,k}_{s,t}$ and $r^k_{i,j}$ are unknown variables. Eq. (16) is an expression commonly used in network flow theory to denote the connectivity constraint. For a backup path $B^k_{i,j}$, Eq. (16) means that the logical links selected for $B^k_{i,j}$ are required to form a path from node v_i to node v_j. The first sum in Eq. (16) is the number of the logical links on $B^k_{i,j}$ that leave node v_s. The second sum is the number of the logical links on $B^k_{i,j}$ that enter node v_s. Eq. (16) includes three cases. If v_s is the source, the outgoing logical links should be one more than the ingoing logical links. If v_s is the destination, the outgoing logical links should be one less than the ingoing logical links. If v_s is neither the source nor the destination, the two numbers are equal. The constraint in Eq. (17) requires that the overall rerouted traffic load on each logical link does not exceed its usable bandwidth. Eq. (18) specifies that the rerouted traffic load on the backup paths for $e_{i,j}$ does not exceed $l_{i,j}$. Ideally, all the traffic load of $e_{i,j}$ should be rerouted. In some cases, the network does not have enough bandwidth, and hence the overall rerouted traffic load of $e_{i,j}$ may be lower than $l_{i,j}$ as shown in Eq. (19)

$$\text{minimize} \quad D \tag{15}$$

subject to

$$\forall e_{i,j} \in E_L \quad \forall e_{s,t} \in E_L \quad 1 \le k \le N$$

$$\sum_{\forall t: e_{s,t} \in E_L} x^{i,j,k}_{s,t} - \sum_{\forall t: e_{t,s} \in E_L} x^{i,j,k}_{t,s} = \begin{cases} 1 & i = s \\ -1 & i = t \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

$$\sum_{e_{i,j} \in E_L} \sum_{k=1}^{N} x^{i,j,k}_{s,t} r^k_{i,j} \le c_{s,t} - l_{s,t} \tag{17}$$

$$\sum_{k=1}^{N} r^k_{i,j} \le l_{i,j} \tag{18}$$

$$x^{i,j,k}_{s,t} \in \{0,1\}, \quad 0 \le r^k_{i,j} \le l_{i,j}. \tag{19}$$

4.3 An Algorithm

Backup paths are computed by routers. Routers in high speed IP networks need to forward packets very quickly. Using too much CPU time for recovery computation interferes with forwarding the packets which are not affected by failures. The above formulation is a mixed integer nonlinear programming problem which is NP-hard. Therefore, routers cannot compute backup paths by directly solving this problem. Instead, we propose a heuristic-based multi-round algorithm SelectBP to efficiently solve the problem. The basic idea is to select backup paths one by one until there is no usable bandwidth or no logical link can have more backup paths. We use $D_{i,j}$ defined in Eq. (13) as the weight of $e_{i,j}$. In each round, the algorithm SelectBP picks out the logical link $e_{i,j}$ with the largest weight, and then selects a backup path for $e_{i,j}$ and determines the rerouted traffic load. Suppose $e_{i,j}$ already has $k-1$ backup paths. Adding one more backup path reduces traffic disruption $D_{i,j}$ by $\Delta_{i,j}$ shown in Eq. (20). The heuristic of the algorithm SelectBP is that it reduces as much $D_{i,j}$ as possible in each round

$$\Delta_{i,j} = p_{i,j} r^k_{i,j}\left(1 - P\left(B^k_{i,j}|e_{i,j}\right)\right). \tag{20}$$

We develop an algorithm MaxWeightPath to choose $B^k_{i,j}$ and determine $r^k_{i,j}$ to maximize $\Delta_{i,j}$. The basic idea is similar to Dijkstra's algorithm for calculating the shortest path. Starting from node v_i, the algorithm gradually adds logical links to expand the backup path. The rerouted traffic load is the smaller one between the unprotected traffic load of $e_{i,j}$ and the usable bandwidth of the backup path. The algorithm MaxWeightPath operates similar to Dijkstra's algorithm, and thus has the same computational complexity $O((|E_L| + |V_L|)\log(|V_L|))$ as Dijkstra's algorithm. Since the network has $|E_L|$ logical links and each of them can have up to N backup paths, the algorithm SelectBP invokes the algorithm MaxWeightPath at most $N|E_L|$ times. Therefore, the computational complexity of the algorithm SelectBP is $O(N|E_L|(|E_L| + |V_L|)\log(|V_L|))$ in the worst case. The detailed pseudocode of the two algorithms is shown in Appendix B which can be found in the online supplemental material.

5 Performance Evaluation

We evaluate the performance of the proposed approach and compare it with other backup path-based approaches.

5.1 Simulation Setup

5.1.1 Network Topology

The simulation is based on four real ISP networks with both physical and logical topologies, i.e., ChinaNet[2,3], Level3[4],

[2] http://www.chinatelecomusa.com/content_images/NationalFiber_Full.jpg

[3] http://www.chinatelecomusa.com/content_images/ChinaNet_Full.jpg

[4] http://www.level3.com/en/About-Us/~/media/Assets/maps/level_3_network_map.ashx

TABLE 2
Real ISP Topologies Used for Evaluation

Network	Physical topology		Logical topology	
	# Nodes	# Fiber links	# Nodes	# Logical links
ChinaNet	85	141	39	61
Level3	209	230	63	285
Qwest	152	181	49	77
XO	61	71	40	65

Qwest[5], and XO.[6] We use the PoP-level logical topology, where nodes correspond to cities. Table 2 summarizes the physical and logical topologies in the four networks. Due to lack of lightpath information, we build the topology mapping with the method in [20] to minimize the number of fiber links shared by logical links.

5.1.2 IP Layer Configuration

The logical link capacity of Qwest and XO networks is set to the value provided in their logical topologies. For ChinaNet and Level3, we use the method in [24] to assign the logical link capacity. This method assumes that high degree nodes are Level-1 PoP. The logical links between Level-1 PoP have high capacity (10Gb/s) and other logical links have low capacity (2.5 Gb/s), which match the recent ISP case studies [25]. Similar to [17], we use the gravity model to generate synthetic traffic demands. This method assumes that the incoming traffic at a PoP is proportional to the combined capacity of its outgoing links. The average logical link utilization of generated traffic is varied from 5 percent to 40 percent in increments of 5 percent. We use 40 percent as the upper bound because ISPs usually upgrade their infrastructure when the logical link utilization exceeds 40 percent [24], [25]. Finally, we use the method in [26] to optimize the link cost based on the link capacity and traffic demands. Each logical topology adopts the shortest path routing calculated based on the optimized link cost.

5.1.3 Failure Scenarios

The link failure scenarios are set according to the Internet measurement [14]. We randomly choose 2.5 percent logical links to be high failure links. Their failure probability is between 0.5 percent and 0.1 percent and satisfies the power-law distribution with parameter -0.73. For other logical links, their failure probability is between 0.1 percent and 0.01 percent and satisfies the power-law distribution with parameter -1.35. The failure probability of fiber links is between 0.05 percent and 0.01 percent and satisfies the power-law distribution with parameter -1. In the simulation, we use 100 failure probability settings and each simulation is run 1,000 times for each failure probability setting. Each time, we randomly choose one logical link and determine if it is an independent failure or a correlated failure based on the failure probability. If it is a correlated failure, we fail one fiber link which is shared by this logical link and other logical links. The logical links embedded on the failed fiber link all fail.

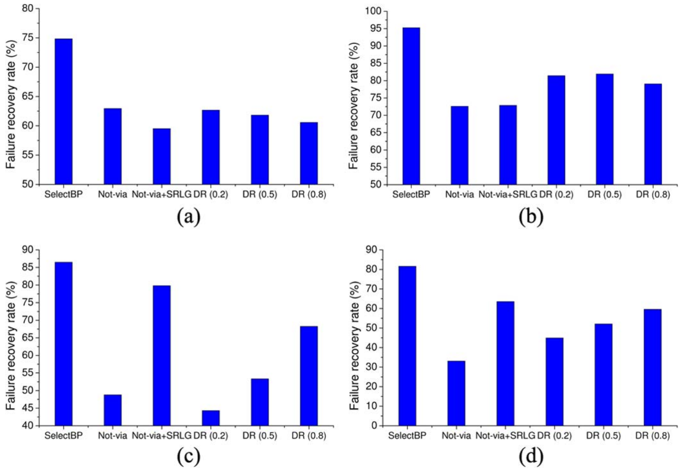


Fig. 3. Average failure recovery rate. (a) ChinaNet. (b) Level3. (c) Qwest. (d) XO.

5.1.4 Algorithms

We compare our algorithms with Not-via [27] and PSRLG-based Diverse Routing (DR) with disjointness constraint [13]. Not-via is an IP fast-rerouting (IPFRR) technique widely deployed in the Internet. DR uses a parameter p to specify the failure probability of logical links when the underlying fiber link fails. We set p to three typical values 0.2, 0.5, and 0.8. In summary, we compare our algorithms with five algorithms as follows.

- *Not-via*: Not-via built on the independent model.
- *Not-via+SRLG*: Not-via built on the SRLG model.
- *DR (0.2)*: DR with its parameter p of 0.2.
- *DR (0.5)*: DR with its parameter p of 0.5.
- *DR (0.8)*: DR with its parameter p of 0.8.

5.2 Reliability of Backup Paths

We first investigate the reliability of the backup paths. In our algorithm SelectBP, the logical link capacity is set to infinity and the parameter N is 1. Hence, the algorithm chooses the backup path with the lowest failure probability for each logical link. It is possible that a logical link may not have a backup path, because the network becomes two connected components after removing the logical link. This may happen to all backup path-based protection approaches. In a test case, if a failed logical link has a live backup path, this logical link failure can be recovered. The *failure recovery rate* is the percentage of recovered logical link failures and is used as a performance metric.

The average failure recovery rate across 100,000 test cases is shown in Fig. 3. We highlight three features. First, our algorithm SelectBP outperforms the other five algorithms in all four networks. This shows that the PCF model is effective for finding reliable backup paths. Not-via ignores the correlation between logical link failures, and thus backup paths may traverse some failed logical links. Not-via+SRLG may remove some useful logical links and even disconnect the topology. Consequently, some logical links may not have backup paths. Second, unlike SelectBP, the performance of the other five algorithms is not consistent across the four networks. For example,

[5] http://www.qwest.com/largebusiness/enterprisesolutions/networkMaps/
[6] http://www.xo.com/about/network/Pages/maps.aspx

Not-via+SRLG is the second best in Qwest and XO, while it is the worst in ChinaNet. Similarly, DR (0.8) is better than DR (0.2) and DR (0.5) in Qwest and XO, but it is worse than them in ChinaNet and Level3. Third, the parameter p strongly affects the performance of DR, and it is difficult to choose an appropriate p to achieve good performance in all networks.

The overall result for the four networks is shown in Fig. 4. On average, the reliability of the backup paths chosen by our approach is at least 18 percent higher than that achieved by the best of the other five algorithms, which varies from scenario to scenario.

5.3 Routing Disruption and Logical Link Overload

Next, we consider the traffic load and bandwidth constraint. Parameter N in the algorithm SelectBP is varied from 1 to 5. We define the following two metrics for measuring the benefit and negative impact.

- *Routing disruption*: For a failed logical link $e_{i,j}$, if a backup path does not contain any failed or overloaded logical link, the traffic rerouted by it is recovered. Suppose the overall traffic load of failed logical links is T and the recovered traffic load is T_r, the routing disruption is defined as $\frac{T-T_r}{T}$. The optimal value (0 percent) means that no traffic is disrupted by failures.
- *Overload rate*: In a test case, we count the logical links traversed by the rerouted traffic and denote this number as L. We also count the overloaded ones among them. A logical link is overloaded if its capacity is smaller than the traffic load on it, including its own traffic and the rerouted traffic. Suppose there are L_o overloaded logical links. The overload rate is defined as $\frac{L_o}{L}$, and it quantifies the negative impact caused by the rerouted traffic.

The average routing disruption is shown in Fig. 5. We highlight four important aspects of the simulation results. First, using more backup paths can reduce the routing disruption, especially when the logical link utilization is high. The performance of SelectBP does not vary much when $N \geq 2$ and thus we only show the result when $N = 1$ and $N = 2$. It means that two backup paths should be adequate for protecting a logical link. Second, SelectBP outperforms the other five algorithms under different logical link utilizations in each network. Third, the performance of the other five algorithms is not consistent across the four networks, which is similar to the failure recovery rate in Fig. 3. Our approach is better than the other five in adapting to different networks and logical link utilizations. Fourth, the routing disruption increases as the logical link utilization increases due to lack of usable bandwidth. For example, if the logical link utilization is 25 percent, the usable bandwidth is 3 times the traffic load. However, when it increases to 40 percent, the usable bandwidth decreases to 1.5 times the traffic load. A small increase in the traffic load makes rerouting much more difficult.

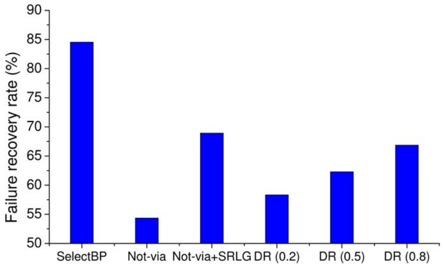


Fig. 4. Overall failure recovery rate for the four networks.

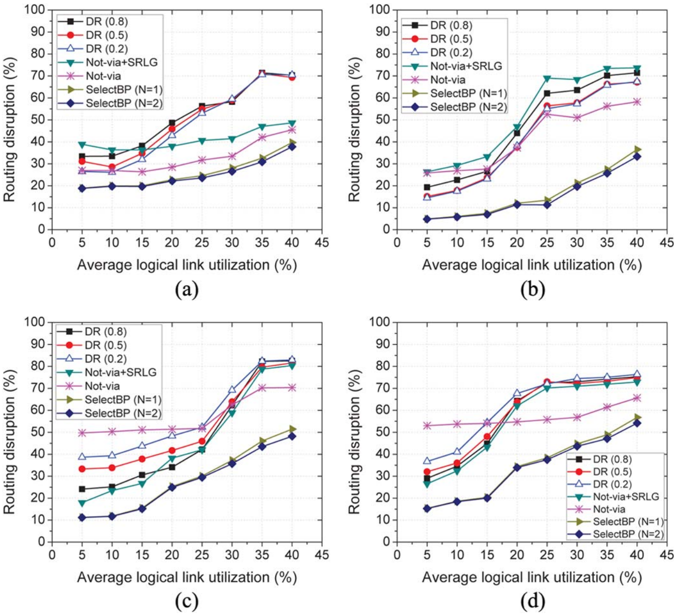


Fig. 5. Average routing disruption under different logical link utilizations. (a) ChinaNet. (b) Level3. (c) Qwest. (d) XO.

The overall result for the four networks is shown in Fig. 6. On average, the routing disruption of our approach is at least 22 percent lower that that of the other five algorithms.

Next, we evaluate the overload rate under different logical link utilizations, which is shown in Fig. 7. SelectBP avoids logical link overload with two techniques, i.e., using logical links with usable bandwidth and controlling the rerouted traffic load. The other five algorithms may have quite high overload rate when the logical link utilization is above 20 percent.

We also measure the maximum logical link utilization on recovery paths, and show the result when the average logical link utilization is 40 percent in Table 3. By considering the bandwidth constraint, the maximum

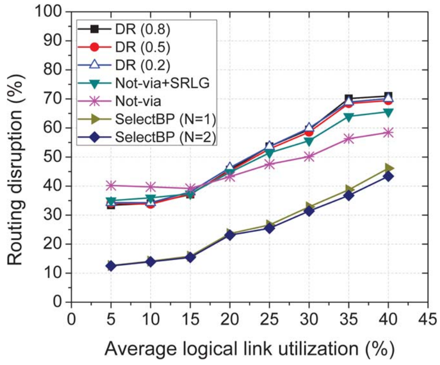
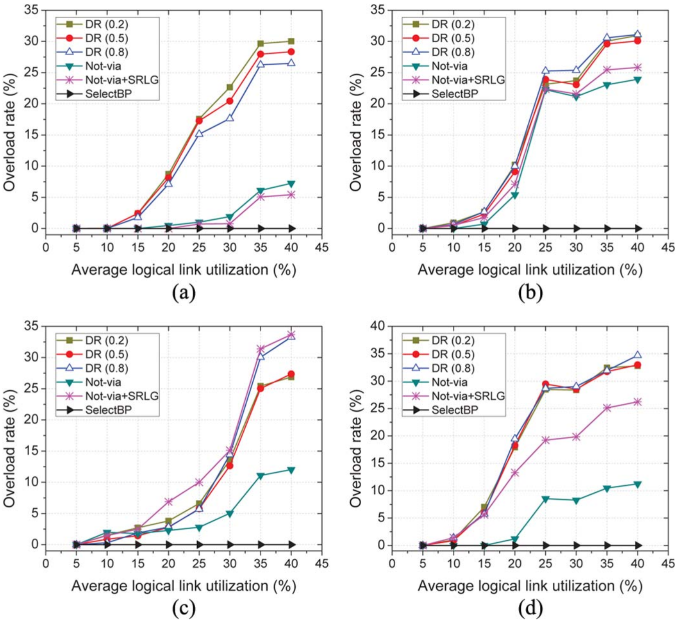


Fig. 6. Overall routing disruption for the four networks.



(a)

(b)

(c)

(d)

Fig. 7. Overload rate under different logical link utilizations. (a) ChinaNet. (b) Level3. (c) Qwest. (d) XO.

logical link utilization in SelectBP is 100 percent, which means that SelectBP fully utilizes the bandwidth and does not cause logical link overload. The other five algorithms do not consider the bandwidth constraint, and hence some logical links may be used by many backup paths at the same time. As a result, the maximum logical link utilization in these algorithms is quite high.

6 Related Work

There are three categories of existing works that are related to our approach.

6.1 Backup Path-Based IP Link Protection

Most prior works consider backup path selection as a connectivity problem and mainly focus on finding backup paths to bypass the failed IP links [4], [7], [8], [9], [13], and [27]. However, they ignore the fact that a backup path may not have enough bandwidth. Consequently, the rerouted traffic may cause severe link overload on an IP backbone as observed by Iyer *et al.* [16]. A recent work [10] addresses the link overload problem in the backup path selection, but it aims at minimizing the bandwidth allocated to backup paths rather than minimizing routing disruption. All these methods use IP layer information for backup path selection, consider logical link failures as independent events, and select one backup path for each logical link. Different from these methods, we develop a PCF model to reflect the probabilistic correlation between logical link failures, and split the rerouted traffic onto multiple backup paths to minimize routing disruption and avoid link overload.

6.2 Correlation between the Logical and Physical Topologies

Some works on IP-over-WDM networks consider the correlation between the logical and physical topologies. Most of them focus on the survivable routing problem [20], [21], [28], and [29], i.e., building the mapping between logical and fiber links to minimize the impact of fiber link failures on logical links. Lee *et al.* [22], [23] showed that the reliability of IP layer is strongly affected by the topology mapping. However, these works do not address the problem of selecting backup paths to protect IP links. Moreover, they do not model the correlation between logical link failures as is done in our PCF model. Cui *et al.* [30] considered correlated failures in backup path allocation for overlay networks. However, they only use overlay layer information, while our approach is based on a cross-layer design. Moreover, they aim at finding reliable backup paths; whereas our objective is to minimize routing disruption. A preliminary version [31] of the paper also considers the topology mapping, but it is different in two aspects. First, the PCF model considers both independent and correlated logical link failures, whereas the model in [31] only considers correlated failures. Second, each logical link is protected by multiple backup paths in this paper, but protected by single backup path in [31].

TABLE 3
Maximum Logical Link Utilization When the Average Logical Link Utilization is 40 Percent

Algorithm	ChinaNet	Level3	Qwest	XO
SelectBP	100	100	100	100
Not-via	232.1	156.8	305.9	145.7
Not-via+SRLG	194.8	203.2	237.8	259.1
DR (0.2)	292.2	294.6	307.7	291.5
DR (0.5)	286.7	343.9	314.8	297.7
DR (0.8)	301.2	281.1	318.3	279.6

6.3 Multipath Routing and Bandwidth Allocation

Quality-of-Service (QoS) routing protocols [32], [33], [34], and [35] use multiple paths between a source-destination pair to achieve traffic engineering goals, e.g., minimizing the maximal link utilization. Kodialam *et al.* [36] proposed an algorithm for dynamic routing of bandwidth guaranteed tunnels. However, they do not consider the correlation between logical link failures. There are few recovery approaches that are built on multiple recovery paths. The approach in [37] aims at minimizing the bandwidth reserved for backup paths. It only uses IP layer information for backup path selection and assumes that the network has a single logical link failure. R3 [17] reroutes traffic with multiple paths and the method in [38] jointly addresses failure recovery and traffic engineering in multipath routing. They focus on traffic engineering goals rather than minimizing routing disruption. Moreover, they ignore the correlation between logical link failures and consider backup paths to have the same reliability.

7 Conclusion

The commonly used independent and SRLG models ignore the correlation between the optical and IP layer topologies. As a result, they do not accurately reflect the correlation between logical link failures and may not select reliable backup paths. We propose a cross-layer approach for minimizing routing disruption caused by IP link failures.

We develop a probabilistically correlated failure (PCF) model to quantify the impact of IP link failure on the reliability of backup paths. With this model, we propose an algorithm to minimize the routing disruption by choosing multiple reliable backup paths to protect each IP link. The proposed approach ensures that the rerouted traffic does not cause logical link overload, even when multiple logical links fail simultaneously. We evaluate our approach using real ISP networks with both optical and IP layer topologies. Experimental results show that two backup paths are adequate for protecting a logical link. Compared with existing works, the backup paths selected by our method are at least 18 percent more reliable and the routing disruption is reduced by at least 22 percent. Moreover, the proposed approach prevents logical link overload caused by the rerouted traffic.

REFERENCES

[1] Q. Zheng, G. Cao, T.L. Porta, and A. Swami, "Optimal Recovery from Large-Scale Failures in IP Networks," in *Proc. IEEE ICDCS*, 2012, pp. 295-304.

[2] A. Bremler-Barr, Y. Afek, H. Kaplan, E. Cohen, and M. Merritt, "Restoration by Path Concatenation: Fast Recovery of MPLs Paths," in *Proc. ACM PODC*, 2001, pp. 43-52.

[3] V. Sharma and F. Hellstrand, *Framework for MPLS-Based Recovery,* RFC 3469, 2003.

[4] M. Shand and S. Bryant, *IP Fast Reroute Framework,* RFC5714, Jan. 2010.

[5] P. Francois, C. Filsfils, J. Evans, and O. Bonaventure, "Achieving Sub-Second IGP Convergence in Large IP Networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 3, pp. 35-44, July 2005.

[6] F. Giroire, A. Nucci, N. Taft, and C. Diot, "Increasing the Robustness of IP Backbones in the Absence of Optical Level Protection," in *Proc. IEEE INFOCOM*, 2003, pp. 1-11.

[7] A. Kvalbein, A.F. Hansen, T. Cicic, S. Gjessing, and O. Lysne, "Fast IP Network Recovery Using Multiple Routing Configurations," in *Proc. IEEE INFOCOM*, 2006, pp. 1-11.

[8] S. Kini, S. Ramasubramanian, A. Kvalbein, and A.F. Hansen, "Fast Recovery from Dual Link Failures in IP Networks," in *Proc. IEEE INFOCOM*, 2009, pp. 1368-1376.

[9] M. Hou, D. Wang, M. Xu, and J. Yang, "Selective Protection: A Cost-Efficient Backup Scheme for Link State Routing," in *Proc. IEEE ICDCS*, 2009, pp. 68-75.

[10] M. Johnston, H.-W. Lee, and E. Modiano, "A Robust Optimization Approach to Backup Network Design with Random Failures," in *Proc. IEEE INFOCOM*, 2011, pp. 1512-1520.

[11] E. Oki, N. Matsuura, K. Shiomoto, and N. Yamanaka, "A Disjoint Path Selection Scheme with Shared Risk Link Groups in GMPLS Networks," *IEEE Commun. Lett.*, vol. 6, no. 9, pp. 406-408, Sept. 2002.

[12] L. Shen, X. Yang, and B. Ramamurthy, "Shared Risk Link Group (SRLG)-Diverse Path Provisioning Under Hybrid Service Level Agreements in Wavelength-Routed Optical Mesh Networks," *Proc. IEEE/ACM Trans. Netw.*, vol. 13, no. 4, pp. 918-931, Aug. 2005.

[13] H.-W. Lee and E. Modiano, "Diverse Routing in Networks with Probabilistic Failures," in *Proc. IEEE INFOCOM*, 2009, pp. 1035-1043.

[14] A. Markopoulou, G. Iannaccone, S. Bhattacharyya, C.-N. Chuah, Y. Ganjali, and C. Diot, "Characterization of Failures in an Operational IP Backbone Network," *IEEE/ACM Trans. Netw.*, vol. 16, no. 4, pp. 749-762, Aug. 2008.

[15] D. Turner, K. Levchenko, A.C. Snoeren, and S. Savage, "California Fault Lines: Understanding the Causes and Impact of Network Failures," in *Proc. ACM SIGCOMM*, 2010, pp. 315-326.

[16] S. Iyer, S. Bhattacharyya, N. Taft, and C. Diot, "An Approach to Alleviate Link Overload as Observed on an IP Backbone," in *Proc. IEEE INFOCOM*, 2003, pp. 406-416.

[17] Y. Wang, H. Wang, A. Mahimkar, R. Alimi, Y. Zhang, L. Qiu, and Y.R. Yang, "R3: Resilient Routing Reconfiguration," in *Proc. ACM SIGCOMM*, 2010, pp. 291-302.

[18] Q. Zheng and G. Cao, "Minimizing Probing Cost and Achieving Identifiability in Probe Based Network Link Monitoring," *IEEE Trans. Comput.*, vol. 62, no. 3, pp. 510-523, Mar. 2013.

[19] E. Rosen, A. Viswanathan, and R. Callon, *Multiprotocol Label Switching Architecture,* RFC 3031, Jan. 2001.

[20] E. Modiano and A. Narula-Tam, "Survivable Lightpath Routing: A New Approach to the Design of WDM-Based Networks," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 4, pp. 800-809, May 2002.

[21] A. Todimala and B. Ramamurthy, "A Scalable Approach for Survivable Virtual Topology Routing in Optical WDM Networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 6, pp. 63-69, Aug. 2007.

[22] K. Lee and E. Modiano, "Cross-Layer Survivability in WDM-Based Networks," in *Proc. IEEE INFOCOM*, 2009, pp. 1017-1025.

[23] K. Lee, H.-W. Lee, and E. Modiano, "Reliability in Layered Networks with Random Link Failures," in *Proc. IEEE INFOCOM*, 2010, pp. 1-9.

[24] S. Kandula, D. Katabi, B. Davie, and A. Charny, "Walking the Tightrope: Responsive Yet Stable Traffic Engineering," in *Proc. ACM SIGCOMM*, 2005, pp. 253-264.

[25] J. Guichard, F. le Faucheur, and J.P. Vasseur, *Definitive MPLS Network Design*. Indianapolis, IN, USA: Cisco Press, 2005.

[26] B. Fortz and M. Thorup, "Internet Traffic Engineering by Optimizing OSPF Weights," in *Proc. IEEE INFOCOM*, 2000, pp. 519-528.

[27] S. Bryant, M. Shand, and S. Previdi, *IP Fast Reroute Using Not-via Addresses,* Internet draft, 2011. [Online]. Available: http://tools.ietf.org/html/draft-ietf-rtgwg-ipfrr-notvia-addresses-03

[28] M. Kurant and P. Thiran, "On Survivable Routing of Mesh Topologies in IP-Over-WDM Networks," in *Proc. IEEE INFOCOM*, 2005, pp. 1106-1116.

[29] K. Thulasiraman, M.S. Javed, and G.L. Xue, "Circuits/Cutsets Duality and a Unified Algorithmic Framework for Survivable Logical Topology Design in IP-over-WDM Optical Networks," in *Proc. IEEE INFOCOM*, 2009, pp. 1026-1034.

[30] W. Cui, I. Stoica, and R.H. Katz, "Backup Path Allocation Based on a Correlated Link Failure Probability Model in Overlay Networks," in *Proc. IEEE ICNP*, 2002, pp. 236-245.

[31] Q. Zheng, J. Zhao, and G. Cao, "A Cross-Layer Approach for IP Network Protection," in *Proc. IEEE/IFIP DSN*, 2012, pp. 1-12.

[32] A. Orda and A. Sprintson, "Efficient Algorithms for Computing Disjoint QoS Paths," in *Proc. IEEE INFOCOM*, 2004, pp. 727-738.

[33] Y. Ohara, S. Imahori, and R.V. Meter, "MARA: Maximum Alternative Routing Algorithm," in *Proc. IEEE INFOCOM*, 2009, pp. 298-306.

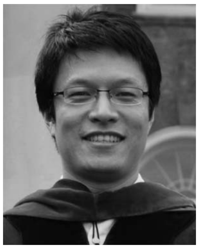
[34] S. Misra, G. Xue, and D. Yang, "Polynomial Time Approximations for Multi-Path Routing with Bandwidth and Delay Constraints," in *Proc. IEEE INFOCOM*, 2009, pp. 558-566.

[35] W. Zhang, J. Tang, C. Wang, and S. de Soysa, "Reliable Adaptive Multipath Provisioning with Bandwidth and Differential Delay Constraints," in *Proc. IEEE INFOCOM*, 2010, pp. 1-9.

[36] M. Kodialam and T.V. Lakshman, "Minimum Interference Routing with Applications to MPLS Traffic Engineering," in *Proc. IEEE INFOCOM*, 2000, pp. 884-893.

[37] R. Banner and A. Orda, "Designing Low-Capacity Backup Networks for Fast Restoration," in *Proc. IEEE INFOCOM*, 2010, pp. 1-9.

[38] M. Suchara, D. Xu, R. Doverspike, D. Johnson, and J. Rexford, "Network Architecture for Joint Failure Recovery and Traffic Engineering," in *Proc. ACM SIGMETRICS*, 2011, pp. 97-108.



Qiang Zheng received the BS degree in computer science from Nankai University, Tianjin, China, in 2004, the ME degree in computer science from Chinese Academy of Sciences, Beijing, China, in 2007, and the PhD degree in computer science and engineering from The Pennsylvania State University, State College, PA, USA, in 2012. His research interests include network failure detection, localization, and fast recovery. He is a Student Member of the IEEE.

Guohong Cao received the BS degree in computer science from Xian Jiaotong University, Shaanxi, China, and the PhD degree in computer science from the Ohio State University, Columbus, OH, USA, in 1999. Since then, he has been with the Department of Computer Science and Engineering at the Pennsylvania State University, State College, PA, USA where he is currently a Professor. He has published more than 200 papers in the areas of wireless networks, wireless security, vehicular networks, wireless sensor networks, cache management, and distributed fault tolerant computing. He has served on the editorial board of *IEEE Transactions on Mobile Computing*, *IEEE Transactions on Wireless Communications*, *IEEE Transactions on Vehicular Technology*, and has served on the organizing and technical program committees of many conferences, including the TPC Chair/Co-Chair of IEEE SRDS'2009, MASS'2010, and INFOCOM'2013. He was a recipient of the NSF CAREER award in 2001. He is a Fellow of the IEEE.

Thomas F. La Porta is the William E. Leonhard Chair Professor in the Computer Science and Engineering Department at Penn State, PA, USA. He is the Director of the Networking and Security Research Center at Penn State. Prior to joining Penn State, Dr. La Porta was with Bell Laboratories since 1986. He was the Director of the Mobile Networking Research Department in Bell Laboratories, Lucent Technologies where he led various projects in wireless and mobile networking. He is a Bell Labs Fellow and he received the Bell Labs Distinguished Technical Staff Award in 1996. He also won a Thomas Alva Edison Patent Awards in 2005 and 2009. Dr. La Porta was the founding Editor-in-Chief of the *IEEE Transactions on Mobile Computing*. He served as Editor-in-Chief of *IEEE Personal Communications Magazine*. He was the Director of Magazines for the IEEE Communications Society and was on its Board of Governors for three years. He has published numerous papers and holds 35 patents. He was an adjunct member of faculty at Columbia University for 7 years where he taught courses on mobile networking and protocol design. He is a Fellow of the IEEE.

Ananthram Swami received the BTech degree from IIT-Bombay, Maharashtra, India, the MS degree from Rice University, Houston, TX, USA and the PhD degree from the University of Southern California (USC), Los Angeles, CA, USA, all in electrical engineering. He is with the U.S. Army Research Laboratory (ARL) as the Army's Senior Research Scientist (ST) for Network Science. He is an ARL Fellow. He has held positions with Unocal Corporation, the University of Southern California (USC), CS-3, and Malgudi Systems. He was a Statistical Consultant to the California Lottery, developed a MATLAB-based toolbox for non-Gaussian signal processing, and has held visiting faculty positions at INP, Toulouse. His research interests are in the broad area of network science, with an emphasis on applications in composite tactical networks. He is a Fellow of the IEEE.