A Cross-Layer Approach for IP Network Protection

Qiang Zheng, Jing Zhao, and Guohong Cao
Department of Computer Science and Engineering
The Pennsylvania State University
E-mail: {quz103, juz139, gcao}@cse.psu.edu

Abstract—Backup paths are widely used to protect IP links from failures. Existing solutions such as the commonly used independent and Shared Risk Link Group models 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 IP link protection. We develop a correlated failure probability (CFP) model to quantify the impact of an IP link failure on the reliability of backup paths. With the CFP model, we propose two algorithms for selecting backup paths. The first algorithm focuses on choosing the backup paths with minimum failure probability. The second algorithm further considers the bandwidth constraint and aims at minimizing the traffic disruption caused by failures. It also ensures that the rerouted traffic load on each IP link does not exceed the usable bandwidth to avoid interfering with the normal traffic. Simulations based on real ISP networks show that our approach can choose backup paths that are more reliable and achieve better protection.

Index Terms—IP link, protection, cross-layer, backup path, model

I. INTRODUCTION

The Internet has become a significant and widely used infrastructure for many communication services. Internet robustness becomes more and more important due to the broad deployment of applications with strict demands on network reliability and availability, such as financial transactions, online games, and multimedia services. However, IP link failures are fairly common in the Internet [1]. Quickly recovering routing from IP link failures is crucial to enhancing Internet reliability and availability, and has received much attention in recent years.

Currently, backup path-based protection [2], [3] is widely used by Internet Service Providers (ISPs) to recover intra-domain routing from IP link failures. In this approach, backup paths are precomputed, configured, and stored in routers. When an IP link failure is detected, traffic originally traversing the IP link is immediately switched to the backup path. Through this, the routing disruption duration can be reduced to the failure detection time which is typically less than 10 milliseconds [4].

Selecting backup paths is a critical problem in backup path-based protection. Existing approaches mainly focus on choosing reliable backup paths to protect IP links. 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 the IP links.

Most IP backbone networks are built on the Wavelength Division Multiplexing (WDM) infrastructure [5]. 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. A logical link may consist of multiple fiber links and a fiber link may be shared by multiple logical links. When a fiber link fails, all the logical links embedded on it fail simultaneously. Fig. 1 shows an example of the topology mapping. The logical topology in Fig. 1(a) is embedded on the physical topology shown in Fig. 1(b), where 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. 1(c). In the past, the optical layer had protection mechanisms to deal with fiber link failures. Such protection usually requires hardware redundancy and is quite expensive [6]. Therefore, most ISPs removed optical layer protection and use backup paths in the IP layer to restore the connectivity [5]. As a result, the model of logical link failures used for backup path selection should reflect the correlation between the logical and physical topologies.

In prior works, logical link failures are considered as independent events [7]–[11] or modeled as a Shard Risk Link Group (SRLG[1]) [12]–[14]. However, both models have limitations. First, logical link failures are not independent because of the topology mapping. Second, sharing fiber links does not imply logical links in the same SRLG must fail simultaneously. Fig. 1 shows an example. If the failure of $e_{1,4}$ is caused by $f_{1,5}$, $e_{1,2}$ also fails. If it is caused by $f_{4,5}$, $e_{1,2}$ may be live. Logical links $e_{1,2}$ and $e_{1,4}$ are in the same SRLG, but they need not fail simultaneously. Their failures are correlated with a certain probability. This feature cannot be modeled by the traditional independent and SRLG models, and has not been studied in backup path selection.

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. Consequently, the rerouted traffic load on some logical links may exceed their usable bandwidth, and thus cause logical link overload. As Iyer *et al.* observed on a major IP backbone [15], most logical

---

[1]A SRLG is a set of logical links that share the same risk such as the fiber link failure. If an IP link fails, all the IP links within the same SRLG are considered as failed.
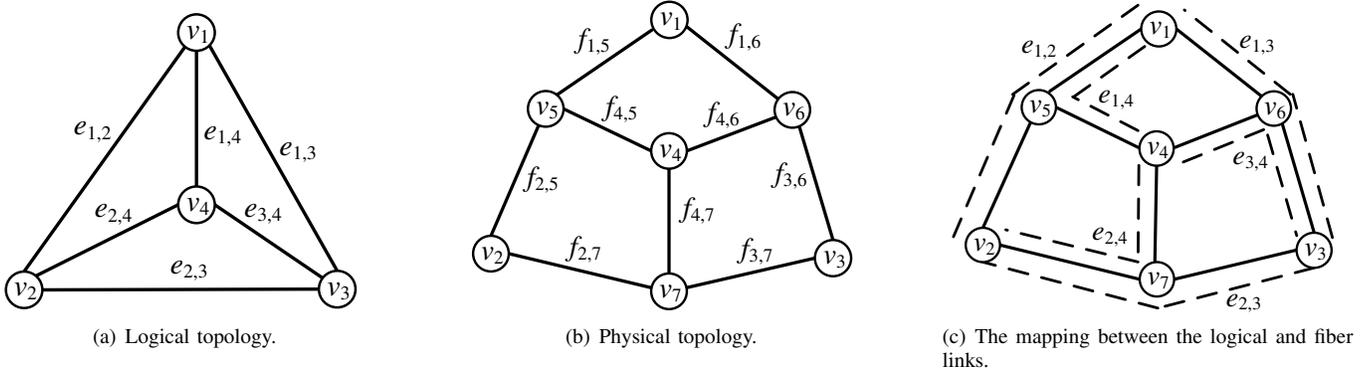
Fig. 1. An example of the mapping between the logical and physical topologies.

link overload 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 [16]. Therefore, backup paths should be carefully chosen to avoid logical link overload.

We propose a cross-layer approach to protect IP links with backup paths. We develop a correlated failure probability (CFP) model based on the topology mapping and failure probability of fiber links. It calculates the failure probability of fiber links, logical links, and backup paths under the condition that a logical link fails. With this model, we propose two algorithms for selecting backup paths. The first algorithm focuses on reliability and aims at choosing the backup paths with minimum failure probability. The second algorithm further considers the bandwidth constraint and aims at choosing backup paths to minimize the traffic disruption caused by failures. Furthermore, it controls the rerouted traffic load to avoid logical link overload, even when multiple logical links fail simultaneously.

Our cross-layer approach is different from prior works in two aspects. First, it is based on a cross-layer design, which considers the correlation between the logical and physical topologies. The proposed model can reflect the correlation between logical link failures. Second, our approach addresses reliability and avoids logical link overload. It guarantees that the rerouted traffic load does not exceed the usable bandwidth of logical links. We evaluate the proposed approach using real ISP networks with both optical and IP layer topologies. Compared with existing works, the backup paths selected by our approach are much more reliable. Moreover, the proposed approach achieves higher recovery rate without causing logical link overload.

The rest of this paper is organized as follows. Section II presents the preliminaries of our work. Section III presents the CFP model. Section IV describes the problem of backup path selection, including the problem formulation and the algorithms. Section V presents the performance evaluation and Section VI reviews related work. Finally, Section VII concludes the paper.

II. PRELIMINARIES

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

A. Backup Path-Based IP Link Protection

Backup path-based IP link protection is widely used in intra-domain routing. In this approach, a router precomputes a backup path for each of its logical links. When a logical link fails, only the two routers connected by it can detect the failure. Upon detecting the failure, the router immediately switches the traffic originally sent on this logical link onto the corresponding backup path. The routing protocol has its mechanism to disseminate failure information and trigger the convergence process. Since most IP link failures are temporary and last for few seconds, rapidly triggering the convergence may cause route flapping and make networks more unstable [17]. Therefore, routers need to wait for several seconds before disseminating failure information.

The backup path is used to reroute traffic until the routing protocol converges to a new network topology. Then, routers compute backup paths based on the new network topology. In Fig. 1(a), suppose that node v_1 adopts $v_1 \rightarrow v_2 \rightarrow v_4$ as the backup path for $e_{1,4}$. When v_1 detects that $e_{1,4}$ fails, it forwards the packets towards v_4 along $v_1 \rightarrow v_2 \rightarrow v_4$. After the routing protocol converges to a new network topology, v_1 has a new shortest path towards v_4, and thus it stops using the backup path. Backup paths can be implemented with Multi-Protocol Label Switching [18] which is widely supported in the current Internet. Each backup path is configured as a Label-Switched Path (LSP) and routers can control the rerouted traffic load on a backup path.

B. Model of IP-over-WDM Networks

The IP-over-WDM network under study has a logical topology and a physical topology, which are commonly modeled as two undirected graphs [19], [20], [22]. 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 node $v_i \in V_P$ to node $v_j \in V_P$ is denoted by $f_{i,j}$. 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 node $v_m \in V_L$ to node $v_n \in V_L$ is denoted by $e_{m,n}$. 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 change frequently. When the network administrator adjusts the topology mapping, the topology mapping information at routers can also be updated.

TABLE I
TABLE OF NOTATIONS

Symbols	Meaning
	Section II-B
$G_P = (V_P, F_P)$	physical topology
$G_L = (V_L, E_L)$	logical topology
$f_{i,j}$	fiber link from node $v_i \in V_P$ to node $v_j \in V_P$
$e_{m,n}$	logical link from node $v_m \in V_L$ to node $v_n \in V_L$
	Section III
$p_{i,j}$	failure probability of $f_{i,j}$, where $p_{i,j} \in [0,1)$
$a_{i,j}^{m,n}$	1 if $e_{m,n}$ is embedded on $f_{i,j}$, 0 otherwise
$P(e_{m,n})$	failure probability of $e_{m,n}$
$S_{g,h}^{m,n}$	set of fiber links shared by $e_{g,h}$ and $e_{m,n}$
$P(f_{i,j}|e_{m,n})$	failure probability of $f_{i,j}$ when $e_{m,n}$ fails
$P(e_{g,h}|e_{m,n})$	failure probability of $e_{g,h}$ when $e_{m,n}$ fails
$B_{m,n}$	backup path for logical link $e_{m,n}$
$x_{g,h}^{m,n}$	1 if $B_{m,n}$ traverses $e_{g,h}$, 0 otherwise
$F_{m,n}$	set of fiber links traversed by $B_{m,n}$
$S_B^{m,n}$	set of fiber links shared by $B_{m,n}$ and $e_{m,n}$
$P(B_{m,n}|e_{m,n})$	failure probability of $B_{m,n}$ when $e_{m,n}$ fails
	Section IV-B
$c_{m,n}$	capacity of $e_{m,n}$
$l_{m,n}$	traffic load on $e_{m,n}$ under normal conditions
$r_{m,n}$	rerouted traffic load on $B_{m,n}$
$D_{m,n}$	traffic disruption of $e_{m,n}$

III. CORRELATED FAILURE PROBABILITY MODEL

This section describes the correlated failure probability (CFP) model. We first provide some motivation and then present the details of this model. Finally, we use an example to show how the model works and why it is different from the traditional independent and SRLG models.

A. Motivation

Due to the topology mapping, a logical link $e_{m,n}$ may share fiber links with other logical links. When $e_{m,n}$ fails, routers cannot determine which fiber links cause the failure, and thus cannot know which logical links fail with $e_{m,n}$. Therefore, the failure of $e_{m,n}$ implies that the logical links sharing at least one fiber link with $e_{m,n}$ may also fail with a certain probability. This correlation cannot be modeled by the traditional independent and SRLG models as a *none-or-all* relation. The independent model considers that logical links only have independent failures, whereas the SRLG model considers that logical links only have correlated failures.

We develop a CFP model to deal with the correlation 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 this model, we can choose reliable backup paths to protect logical links.

B. The CFP Model

The CFP model is built on two kinds of information, i.e., the topology mapping and the failure probability of fiber links, both of which are already gathered by ISPs. ISPs configure their topology mapping, and thus they have this information. The failure probability of fiber links can be obtained with the Internet measurement approach [23] deployed at the optical layer, which can detect fiber link failures through SONET alarms. ISPs also maintain failure information, because they monitor the optical layer of their networks.

A key observation is that the backup path for logical link $e_{m,n}$ is used only when $e_{m,n}$ fails. Therefore, the failure probability of the backup path should be computed under the condition that $e_{m,n}$ fails. To compute the failure probability of backup paths, we first compute the failure probability of fiber links and then that of logical links.

Unlike logical links, most fiber link failures are independent [14], [22]. We assume that a fiber link $f_{i,j}$ fails independently with probability $p_{i,j} \in [0,1)$. In practice, we may obtain $p_{i,j}$ based on previous fiber link failures. Let $a_{i,j}^{m,n}$ express the mapping between logical link $e_{m,n}$ and fiber link $f_{i,j}$.

$$a_{i,j}^{m,n} = \begin{cases} 1 & \text{if } e_{m,n} \text{ is embedded on } f_{i,j} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

When a fiber link fails, the logical links embedded on it all fail. Let $P(e_{m,n})$ denote the failure probability of logical link $e_{m,n}$, which is computed by Eq. (2).

$$P(e_{m,n}) = 1 - \prod_{f_{i,j} \in F_P} (1 - a_{i,j}^{m,n} p_{i,j}) \quad (2)$$

Note that $P(e_{m,n})$ is the failure probability of $e_{m,n}$, but it cannot reveal the correlation between logical link failures. Since logical links are embedded on fiber links, we first investigate how a logical link failure affects the failure probability of fiber links. Let $P(f_{i,j}|e_{m,n})$ be the failure probability of $f_{i,j}$ under the condition that $e_{m,n}$ fails. We only need to deal with the fiber link $f_{i,j}$ with failure probability $p_{i,j} > 0$. There are two cases.

- Case 1: $e_{m,n}$ is not embedded on $f_{i,j}$. It means that the failure of $e_{m,n}$ is not related with the failure of $f_{i,j}$. Therefore, $P(f_{i,j}|e_{m,n})$ is equal to $p_{i,j}$.
- Case 2: $e_{m,n}$ is embedded on $f_{i,j}$. Because $e_{m,n}$ fails, it indicates that $P(e_{m,n}) \neq 0$. Based on Bayes' theorem, $P(f_{i,j}|e_{m,n})$ is calculated as in Eq. (3). $P(f_{i,j})$ is the failure probability of $f_{i,j}$, which is equal to $p_{i,j}$. $P(e_{m,n}|f_{i,j})$ is the failure probability of $e_{m,n}$ under the condition that $f_{i,j}$ fails. Since $e_{m,n}$ is embedded on $f_{i,j}$, a failure of $f_{i,j}$ must result in a failure of $e_{m,n}$, i.e., $P(e_{m,n}|f_{i,j})$ is 1.

$$P(f_{i,j}|e_{m,n}) = \frac{P(e_{m,n}|f_{i,j})P(f_{i,j})}{P(e_{m,n})}$$
$$= \frac{P(f_{i,j})}{P(e_{m,n})}$$
$$= \frac{p_{i,j}}{P(e_{m,n})} \quad (3)$$

In summary, the failure probability $P(f_{i,j}|e_{m,n})$ is shown in Eq. (4).

$$P(f_{i,j}|e_{m,n}) = \begin{cases} p_{i,j} & a_{i,j}^{m,n} = 0 \\ \frac{p_{i,j}}{P(e_{m,n})} & a_{i,j}^{m,n} = 1 \end{cases} \quad (4)$$

Next, we calculate the failure probability of logical link $e_{g,h}$ under the condition that $e_{m,n}$ fails. Let $P(e_{g,h}|e_{m,n})$ denote this probability. Note that we cannot simply replace $p_{i,j}$ in Eq. (2) with $P(f_{i,j}|e_{m,n})$ to obtain $P(e_{g,h}|e_{m,n})$. In statistical theory, Eq. (2) is correct only when probability $p_{i,j}$ is independent to each other. However, for two fiber links $f_{i,j}$ and $f_{s,t}$, their failure probability $P(f_{i,j}|e_{m,n})$ and $P(f_{s,t}|e_{m,n})$ may not be independent. Fig. 1(c) shows one example. Suppose $e_{1,4}$ fails, which indicates that at least one of $f_{1,5}$ and $f_{4,5}$ fails. If $f_{1,5}$ fails, $f_{4,5}$ may be live or fail. If $f_{1,5}$ is live, $f_{4,5}$ must have failed in order for $e_{1,4}$ to fail. When $e_{1,4}$ fails, the status of $f_{1,5}$ is related with the status of $f_{4,5}$. Therefore, $P(f_{1,5}|e_{1,4})$ is not independent to $P(f_{4,5}|e_{1,4})$.

The correlation between $e_{g,h}$ and $e_{m,n}$ is because they share fiber links. To calculate $P(e_{g,h}|e_{m,n})$, we first identify the fiber links shared by $e_{g,h}$ and $e_{m,n}$. Let $S_{g,h}^{m,n}$ be the set of fiber links shared by $e_{g,h}$ and $e_{m,n}$, which is defined by Eq. (5).

$$S_{g,h}^{m,n} = \{f_{i,j}|a_{i,j}^{g,h}a_{i,j}^{m,n} = 1, \forall f_{i,j} \in F_P\} \quad (5)$$

This set of fiber links *fails* if *any* fiber link in the set fails, and it is *live* when every fiber link in the set is live. Let $P(S_{g,h}^{m,n})$ be the probability that set $S_{g,h}^{m,n}$ fails. Since the fiber links in $S_{g,h}^{m,n}$ fail independently, we can calculate $P(S_{g,h}^{m,n})$ with Eq. (6).

$$P(S_{g,h}^{m,n}) = \begin{cases} 1 - \prod_{f_{i,j} \in S_{g,h}^{m,n}}(1 - p_{i,j}) & \text{if } S_{g,h}^{m,n} \neq \emptyset \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

According to the definition of $S_{g,h}^{m,n}$, $e_{m,n}$ is embedded on every fiber link in $S_{g,h}^{m,n}$. Therefore, we can use the method similar to Eq. (4) to compute the failure probability of $S_{g,h}^{m,n}$ under the condition that $e_{m,n}$ fails, which is shown in Eq. (7).

$$P(S_{g,h}^{m,n}|e_{m,n}) = \frac{P(S_{g,h}^{m,n})}{P(e_{m,n})} \quad (7)$$

The fiber links carrying $e_{g,h}$ can be divided into two sets. The first set $S_{g,h}^{m,n}$ contains the fiber links shared by $e_{m,n}$ and $e_{g,h}$, and thus these fiber links are related with the failure of $e_{m,n}$. The second set consists of the fiber links that carry $e_{g,h}$ but do not carry $e_{m,n}$. Any fiber link $f_{i,j}$ in the second set is

not related with the failure of $e_{m,n}$. Hence, $P(f_{i,j}|e_{m,n}) = p_{i,j}$ according to Eq. (4) and $P(f_{i,j}|e_{m,n})$ is independent to $P(S_{g,h}^{m,n}|e_{m,n})$. In summary, $P(e_{g,h}|e_{m,n})$ can be calculated with Eq. (8).

$$P(e_{g,h}|e_{m,n}) = 1 - (1 - P(S_{g,h}^{m,n}|e_{m,n})) \times \prod_{f_{i,j} \in F_P - S_{g,h}^{m,n}}(1 - a_{i,j}^{g,h}p_{i,j}) \quad (8)$$

Finally, we calculate the failure probability of the backup path $B_{m,n}$ for $e_{m,n}$ under the condition that $e_{m,n}$ fails. This probability is denoted by $P(B_{m,n}|e_{m,n})$. We define variable $x_{g,h}^{m,n} \in \{0,1\}$ as follows.

$$x_{g,h}^{m,n} = \begin{cases} 1 & \text{if } B_{m,n} \text{ traverses logical link } e_{g,h} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

In the independent and SRLG models, $P(B_{m,n}|e_{m,n})$ is computed based on the failure probability of logical links as shown in Eq. (10).

$$P(B_{m,n}|e_{m,n}) = 1 - \prod_{e_{g,h} \in E_L}(1 - x_{g,h}^{m,n}P(e_{g,h}|e_{m,n})) \quad (10)$$

This equation enumerates logical links and count the ones traversed by $B_{m,n}$. However, it may count a fiber link multiple times if this fiber link is shared by multiple logical links on $B_{m,n}$. As shown in Fig. 1, path $v_1 \rightarrow v_3 \rightarrow v_4$ traverses fiber link $f_{3,6}$ twice, and thus Eq. (10) counts $f_{3,6}$ twice. Actually, we should count each fiber link only once, because fiber link failures have the same effect on disconnecting $B_{m,n}$. A failure of any underlying fiber link leads to a failure of $B_{m,n}$. Therefore, we compute $P(B_{m,n}|e_{m,n})$ based on fiber links rather than logical links. Similar to the calculation of $P(e_{g,h}|e_{m,n})$, we first determine the fiber links shared by $B_{m,n}$ and $e_{m,n}$. Let $S_B^{m,n}$ be the set of these fiber links, which is defined in Eq. (11).

$$S_B^{m,n} = \{f_{i,j}|x_{g,h}^{m,n}a_{i,j}^{g,h}a_{i,j}^{m,n} = 1, \forall e_{g,h} \in E_L, \forall f_{i,j} \in F_P\} \quad (11)$$

Like set $S_{g,h}^{m,n}$, set $S_B^{m,n}$ *fails* if any fiber link in it fails, and it is live when every fiber link in it is live. Then, we calculate the failure probability $P(S_B^{m,n})$ and the conditional failure probability $P(S_B^{m,n}|e_{m,n})$ similar to $P(S_{g,h}^{m,n})$ and $P(S_{g,h}^{m,n}|e_{m,n})$. The expressions are similar to Eq. (6) and Eq. (7) (by replacing $S_{g,h}^{m,n}$ with $S_B^{m,n}$), and hence we omit them.

Finally, we compute $P(B_{m,n}|e_{m,n})$ with the method similar to the calculation of $P(e_{g,h}|e_{m,n})$ in Eq. (8). To simplify the expression, we introduce notation $F_{m,n}$ to be the set of fiber links traversed by $B_{m,n}$, which is shown in Eq. (12).

$$F_{m,n} = \{f_{i,j}|x_{g,h}^{m,n}a_{i,j}^{g,h} = 1, \forall e_{g,h} \in E_L, \forall f_{i,j} \in F_P\} \quad (12)$$

Note that $S_B^{m,n}$ is a subset of $F_{m,n}$. The fiber links in $F_{m,n} - S_B^{m,n}$ are not related with the failure of $e_{m,n}$, because they do not carry $e_{m,n}$. According to Eq. (4), we have $P(f_{i,j}|e_{m,n}) = p_{i,j}$ for each of these fiber links. In summary, $P(B_{m,n}|e_{m,n})$ is computed by Eq. (13), which has two features. First, each fiber link carrying $B_{m,n}$ is counted once. Second, the fiber links related with the failure of $e_{m,n}$ are separated from the fiber links independent to the failure of $e_{m,n}$. In Section IV-A, we propose an algorithm to choose $B_{m,n}$ with minimum $P(B_{m,n}|e_{m,n})$.

$$P(B_{m,n}|e_{m,n}) = 1 - (1 - P(S_B^{m,n}|e_{m,n})) \times \prod_{f_{i,j} \in F_{m,n} - S_B^{m,n}} (1 - p_{i,j}) \quad (13)$$

C. An Example

We use an example to show how the CFP model works and its difference from the traditional independent and SRLG models. The physical and logical topologies are shown in Fig. 1. The failure probability of fiber links is shown in Fig. 2(a). Suppose v_1 needs to choose a backup path for $e_{1,4}$. First, Eq. (2) calculates the failure probability $P(e_{m,n})$ for each logical link $e_{m,n}$. The result is shown in Fig. 2(b). For example, $P(e_{1,4}) = 1 - (1-0.02)(1-0.01) = 0.03$. Then, we use Eq. (4) to calculate the failure probability $P(f_{i,j}|e_{1,4})$. The result is shown in Fig. 2(c). Since $e_{1,4}$ is embedded on $f_{1,5}$ and $f_{4,5}$, $P(f_{1,5}|e_{1,4})$ is different from $p_{1,5}$ and $P(f_{4,5}|e_{1,4})$ is different from $p_{4,5}$. For example, $P(f_{1,5}|e_{1,4}) = \frac{p_{1,5}}{P(e_{1,4})} = \frac{0.02}{0.03} = 0.67$. Next, we use Eq. (5)–Eq. (8) to compute $P(e_{m,n}|e_{1,4})$ and show the result in Fig. 2(d). For $e_{1,2}$, set $S_{1,2}^{1,4}$ contains $f_{1,5}$, because $e_{1,4}$ and $e_{1,2}$ only share $f_{1,5}$. Hence, the failure probability $P(S_{1,2}^{1,4}) = 1 - (1 - p_{1,5}) = 0.02$ and $P(S_{1,2}^{1,4}|e_{1,4}) = \frac{P(S_{1,2}^{1,4})}{P(e_{1,4})} = \frac{0.02}{0.03} = \frac{2}{3}$. Finally, $P(e_{1,2}|e_{1,4}) = 1 - (1 - P(S_{1,2}^{1,4}|e_{1,4}))(1 - p_{2,5}) = 1 - (1 - \frac{2}{3})(1 - 0.01) = 0.67$.

TABLE II
THE FAILURE PROBABILITY $P(e_{m,n}|e_{1,4})$ COMPUTED BY THE THREE MODELS

Model name	$e_{1,2}$	$e_{1,3}$	$e_{1,4}$	$e_{2,3}$	$e_{2,4}$	$e_{3,4}$
Independent	0.03	0.03	1	0.02	0.02	0.03
SRLG	1	0.03	1	0.02	0.02	0.03
CFP	0.67	0.03	1	0.02	0.02	0.03

Table II summarizes the failure probability $P(e_{m,n}|e_{1,4})$ computed by the two existing models and our model. The models lead to different result for $e_{1,2}$, because only $e_{1,2}$ shares a fiber link with $e_{1,4}$. Due to the omission of topology correlation, the independent model considers that logical link failures are independent, and hence it underestimates $P(e_{1,2}|e_{1,4})$. The SRLG model overestimates $P(e_{1,2}|e_{1,4})$, because it considers that the logical links sharing fiber links with $e_{1,4}$ fail simultaneously.

Finally, we compute the failure probability $P(B_{1,4}|e_{1,4})$. The backup path $B_{1,4}$ can be chosen from four loop-free paths as shown in the first column of Table III. The independent
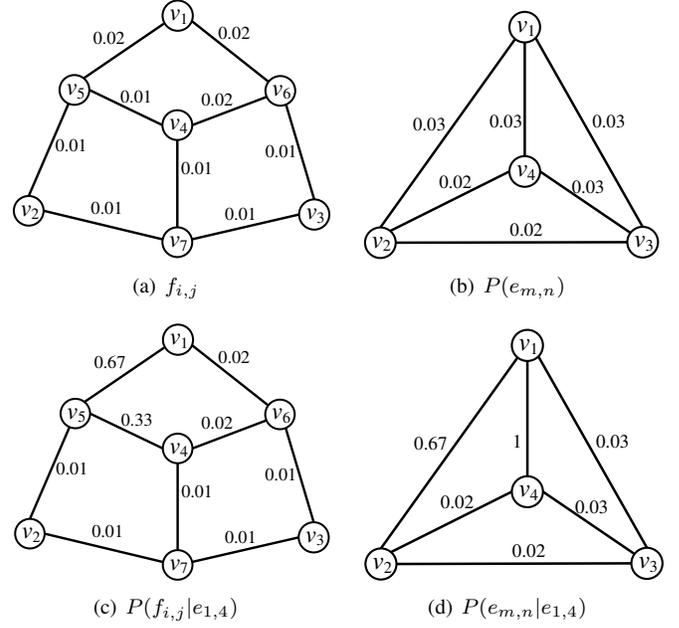

Fig. 2. An example of calculating failure probability with the CFP model.

TABLE III
THE FAILURE PROBABILITY $P(B_{1,4}|e_{1,4})$ COMPUTED BY THE THREE MODELS

Backup path	Independent	SRLG	CFP
$v_1 \to v_2 \to v_4$	0.05	1	0.68
$v_1 \to v_3 \to v_4$	0.06	0.06	0.05
$v_1 \to v_2 \to v_3 \to v_4$	0.08	1	0.69
$v_1 \to v_3 \to v_2 \to v_4$	0.07	0.07	0.06

and SRLG models compute $P(B_{1,4}|e_{1,4})$ with Eq. (10) based on the results in Table II. The result is shown in the second and third columns. We use $v_1 \to v_2 \to v_4$ as an example to show how our model computes $P(B_{1,4}|e_{1,4})$. The set $F_{1,4}$ consists of four fiber links $f_{1,5}$, $f_{2,5}$, $f_{2,7}$, and $f_{4,7}$. Since this path shares $f_{1,5}$ with $e_{1,4}$, the set $S_{1,4}^B$ contains $f_{1,5}$. Hence, $P(S_{1,4}^B) = p_{1,5} = 0.02$ and $P(S_{1,4}^B|e_{1,4}) = \frac{0.02}{0.03} = \frac{2}{3}$. Finally, Eq. (13) calculates $P(B_{1,4}|e_{1,4}) = 1 - (1 - P(S_{1,4}^B|e_{1,4}))(1 - p_{2,5})(1 - p_{2,7})(1 - p_{4,7}) = 1 - (1 - \frac{2}{3})(1 - 0.01)^3 \approx 0.68$.

The independent and SRLG models have two limitations in calculation of $P(B_{1,4}|e_{1,4})$. First, the two models do not accurately reflect the correlation between logical link failures. As a result, the independent model may underestimate $P(B_{1,4}|e_{1,4})$ and the SRLG model may overestimate it as shown by the first and third rows of Table III. Second, they may count a fiber link multiple times. For example, the two models count $f_{3,6}$ twice for $v_1 \to v_3 \to v_4$, and count $f_{2,7}$ twice for $v_1 \to v_3 \to v_2 \to v_4$. However, each fiber link carrying the backup path should be counted once, because they have the same effect on disconnecting the backup path.

IV. BACKUP PATH SELECTION

Based on the CFP model, we propose two algorithms to select backup paths. The first algorithm focuses on reliability and aims at choosing the backup paths with minimum failure

probability. The second algorithm further considers the bandwidth constraint and aims at choosing backup paths to minimize the traffic disruption caused by failures. Furthermore, it controls the rerouted traffic load to prevent causing logical link overload.

A. Reliability Oriented Backup Paths

Reliability oriented backup paths aim at maximizing the reliability of backup paths without considering the traffic load of rerouted traffic. Most prior works focus on this objective. In this problem, backup paths do not compete for bandwidth, and thus we can choose them one by one. For each logical link $e_{m,n}$, the objective is to choose the backup path $B_{m,n}$ with minimum failure probability $P(B_{m,n}|e_{m,n})$. Since $B_{m,n}$ is used only when $e_{m,n}$ fails, the objective is equal to minimizing the probability that $e_{m,n}$ and $B_{m,n}$ fail simultaneously.

This problem can be formally defined as an optimization problem as shown in Eq. (14)–Eq. (16). The objective function Eq. (14) means that we aim at selecting the backup path $B_{m,n}$ with minimum $P(B_{m,n}|e_{m,n})$ to protect logical link $e_{m,n}$. $x_{i,j}^{m,n}$ is defined in Eq. (9) to show if $B_{m,n}$ traverses $e_{i,j}$. The connectivity constraint in Eq. (15) requires that the selected logical links form a path from v_m to v_n.

$$\text{minimize} \quad P(B_{m,n}|e_{m,n}) \quad (14)$$

subject to

$$\sum_{j:e_{i,j}\in E_L} x_{i,j}^{m,n} - \sum_{j:e_{j,i}\in E_L} x_{j,i}^{m,n} = \begin{cases} 1 & i = m \\ -1 & i = n \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

$$\forall e_{i,j} \in E_L \quad x_{i,j}^{m,n} \in \{0,1\} \quad (16)$$

The above formulation is a 0-1 nonlinear programming problem, which is NP-hard in general. Since routers calculate backup paths, the algorithm should be simple and have low computational complexity. Complex algorithms consume much computational resources of the routers, and thus affect normal packet forwarding. Therefore, we propose a greedy algorithm SelectBP as shown in Algorithm 1 to efficiently solve it. The basic idea is similar to Dijkstra's algorithm for calculating the shortest path. Starting from node v_m, the algorithm gradually adds logical links to expand the backup path until finding $B_{m,n}$ with minimum $P(B_{m,n}|e_{m,n})$. By storing $B_{m,i}$ for v_i, constructing $B_{m,j}$ in line 16 can be achieved in $O(1)$ time. Therefore, the computational complexity of Algorithm 1 is $O((|E_L| + |V_L|)\log|V_L|)$.

B. Backup Paths with Bandwidth Constraint

Without considering the rerouted traffic load, reliability oriented backup paths may overload logical links, and thus interfere with normal traffic. To address this problem, we choose backup paths considering the bandwidth constraint. The basic idea is to control the rerouted traffic load and only use the logical links with usable bandwidth. Let $c_{m,n}$ denote the capacity of $e_{m,n}$. Under normal conditions, the

---

Algorithm 1 SelectBP

Procedure:

1: $G_L \leftarrow G_L - e_{m,n}$
2: **for** each node $v_i \in V_L$ **do**
3:    $w(v_i) \leftarrow \infty$ ▷ weight of node v_i, which is the failure probability $P(B_{m,i}|e_{m,n})$ of $B_{m,i}$ from v_m to v_i.
4:    $visit(v_i) \leftarrow 0$ ▷ flag to denote if v_i has been visited
5:    $prev(v_i) \leftarrow NULL$ ▷ previous node of v_i on $B_{m,i}$
6: **end for**
7: $w(v_m) \leftarrow 0$
8: Initialize a priority queue Q, and $Q \leftarrow V_L$
9: **while** $Q \neq \emptyset$ **do**
10:    $v_i \leftarrow$ ExtractMin(Q) ▷ extract the node v_i with minimum weight $w(v_i)$ from Q
11:    **if** $w(v_i) = \infty$ **then**
12:       **break**
13:    **end if**
14:    $visit(v_i) \leftarrow 1$
15:    **for** each neighbor v_j of v_i with $visit(v_j) = 0$ **do**
16:       $B_{m,j} \leftarrow$ the path from v_m to v_j constructed from $prev$
17:       **if** $w(v_j) > P(B_{m,j}|e_{m,n})$ **then**
18:          $w(v_j) \leftarrow P(B_{m,j}|e_{m,n})$
19:          $prev(v_j) \leftarrow v_i$
20:       **end if**
21:    **end for**
22: **end while**
23: Construct $B_{m,n}$ from $prev$

---

traffic load on $e_{m,n}$ is $l_{m,n}$ which satisfies $l_{m,n} < c_{m,n}$. Network administrators configure the link cost to achieve traffic engineering goals, and hence traffic load $l_{m,n}$ is known. The bandwidth of $e_{m,n}$ that can be used for backup paths is $c_{m,n} - l_{m,n}$. Logical link $e_{m,n}$ is *overloaded* if the rerouted traffic load on it exceeds $c_{m,n} - l_{m,n}$.

For a large traffic load $l_{m,n}$, there may not be any backup path which has enough bandwidth to support it. A simple solution is to leave $e_{m,n}$ unprotected to avoid logical link overload. However, all traffic on $e_{m,n}$ is disrupted when $e_{m,n}$ fails. Therefore, we adopt another method, i.e., controlling the rerouted traffic load to protect a part of the traffic. In addition to choosing $B_{m,n}$, we determine the traffic load $r_{m,n}$ for $B_{m,n}$ satisfying $r_{m,n} \leq l_{m,n}$. When $e_{m,n}$ fails, node v_m switches the traffic originally on $e_{m,n}$ onto $B_{m,n}$ and controls the rerouted traffic load to be $r_{m,n}$. The unprotected traffic load is $l_{m,n} - r_{m,n}$, which is disrupted when $e_{m,n}$ fails.

Our objective is to minimize the traffic disruption in the entire network. Based on the CFP model, we define the traffic disruption of $e_{m,n}$ in Eq. (17). When $e_{m,n}$ fails, the unprotected traffic is disrupted. If $B_{m,n}$ also fails, the traffic rerouted on it is disrupted. Therefore, $D_{m,n}$ is the mathematical expectation of disrupted traffic load of $e_{m,n}$.

$$D_{m,n} = P(e_{m,n})(r_{m,n}P(B_{m,n}|e_{m,n}) + l_{m,n} - r_{m,n}) \quad (17)$$

The traffic disruption of the entire network is then defined in Eq. (18), which is the mathematical expectation of traffic disruption in the entire network.

$$D = \sum_{e_{m,n} \in E_L} D_{m,n} \tag{18}$$

Our problem can be formally defined as an optimization problem as shown in Eq. (19)–Eq. (22). For each logical link $e_{m,n}$, the connectivity constraint in Eq. (20) requires that the logical links selected by $B_{m,n}$ form a path from v_m to v_n. The constraint in Eq. (21) requires that the rerouted traffic load on each logical link does not exceed its usable bandwidth. Eq. (22) specifies that the rerouted traffic load $r_{m,n}$ does not exceed $l_{m,n}$. Ideally, all the traffic load of $e_{m,n}$ should be rerouted. In some cases, the network does not have enough bandwidth, and hence the overall rerouted traffic load of $e_{m,n}$ may be lower than $l_{m,n}$.

$$\text{minimize} \quad D \tag{19}$$

subject to

$$\forall v_m, v_n \in V_L$$

$$\sum_{j:e_{i,j} \in E_L} x_{i,j}^{m,n} - \sum_{j:e_{j,i} \in E_L} x_{j,i}^{m,n} = \begin{cases} 1 & i = m \\ -1 & i = n \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

$$\forall e_{i,j} \in E_L \quad \sum_{e_{m,n} \in E_L} x_{i,j}^{m,n} r_{m,n} + l_{i,j} \leq c_{i,j} \tag{21}$$

$$\forall e_{m,n}, e_{i,j} \in E_L \quad x_{i,j}^{m,n} \in \{0,1\}, \quad 0 \leq r_{m,n} \leq l_{m,n} \tag{22}$$

The above formulation is a mixed integer nonlinear programming problem which is NP-hard. We propose a multi-round algorithm SelectBP-BC as shown in Algorithm 2 to efficiently solve it. The basic idea is to choose backup paths one by one. Intuitively, a logical link $e_{m,n}$ with large failure probability $P(e_{m,n})$ and large traffic load $l_{m,n}$ should be protected most, because it is easy to fail and its failure disrupts a large amount of traffic. We use $P(e_{m,n})l_{m,n}$ as the weight of $e_{m,n}$ and deal with logical links in the decreasing order of their weights. The algorithm first sort logical links in decreasing order of their weights (line 1). In each round, the algorithm picks out the unprotected logical link $e_{m,n}$ with the largest weight. It removes $e_{m,n}$ (line 9), and then uses Algorithm 1 to compute $B_{m,n}$ with minimum failure probability (line 10). Next, the algorithm determines the rerouted traffic load $r_{m,n}$ (lines 11–16), which is the smaller one between $l_{m,n}$ and the usable bandwidth of $B_{m,n}$. Finally, it updates the usable bandwidth of logical links and removes the logical links without usable bandwidth (lines 17–22). The computational complexity is $O(|E_L|(|E_L| + |V_L|) \log |V_L|)$, because the algorithm invokes Algorithm 1 for each of the $|E_L|$ logical links.

Algorithm 2 SelectBP-BC

Procedure:

1: Sort logical links in the decreasing order of $P(e_{m,n})l_{m,n}$
2: **for** each logical link $e_{i,j} \in$ **do**
3:     $u_{i,j} \leftarrow c_{i,j} - l_{i,j}$ ▷ the usable bandwidth of $e_{i,j}$
4:     **if** $u_{i,j} = 0$ **then**
5:       $G_L \leftarrow G_L - e_{i,j}$ ▷ remove $e_{i,j}$ from G_L if it does not have usable bandwidth
6:     **end if**
7: **end for**
8: **for** each logical link $e_{m,n} \in E_L$ **do**
9:     $G_{temp} \leftarrow G_L - e_{m,n}$
10:    $B_{m,n} \leftarrow \texttt{SelectBP}$ for $e_{m,n}$ on G_{temp}
11:    $r_{m,n} \leftarrow l_{m,n}$
12:    **for** each logical link $e_{i,j}$ on $B_{m,n}$ **do**
13:      **if** $u_{i,j} < r_{m,n}$ **then**
14:        $r_{m,n} \leftarrow u_{i,j}$
15:      **end if**
16:    **end for**
17:    **for** each logical link $e_{i,j}$ on $B_{m,n}$ **do**
18:      $u_{i,j} \leftarrow u_{i,j} - r_{m,n}$
19:      **if** $u_{i,j} = 0$ **then**
20:        $G_L \leftarrow G_L - e_{i,j}$
21:      **end if**
22:    **end for**
23: **end for**

V. PERFORMANCE EVALUATION

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

A. Simulation Setup

Network topology: The simulation is based on four real ISP networks with both physical and logical topologies, i.e., ChinaNet[2][3], Level3[4], Qwest[5], and XO[6]. We use PoP-level logical topology, where nodes correspond to cities. Table IV summarizes the physical and logical topologies in the four networks. Due to lack of lightpath information, we build the topology mapping with the method used in [19] to minimize the number of fiber links shared by logical links.

TABLE IV
REAL ISP NETWORKS 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

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

IP layer configuration: The logical link capacity of Qwest and XO networks is set to that 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 the 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 (2.5Gb/s), which matches the recent ISP case studies [25]. Similar to [16], [24], 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. 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.

Failure scenarios: Generally, the failure probability of the fiber link is quite low. In the simulation, it is a random number between 1% and 0.01% with the precision of 0.01%. We focus on the scenario of single fiber link failure, which may cause multiple logical link failures. In the simulation, we use 100 failure probability settings. For each setting, the simulation is run 1,000 times and each time one fiber link is selected to fail based on the failure probability. The logical links embedded on the failed fiber link all fail. A fiber link failure is counted as one test case, and thus each simulation has 100,000 test cases.

Algorithms: We compare our algorithms with Not-via [27] and PSRLG-based Diverse Routing (DR) with disjointness constraint [14]. 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.

B. Reliability of Backup Paths

We first investigate the reliability of the backup paths. For our cross-layer approach, we use the algorithm SelectBP proposed in Section IV-A to choose backup paths. 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 the recovered logical link failures, and it is used as the performance metric.

The average failure recovery rate across 100,000 test cases is shown in Fig. 3, and the overall failure recovery rate for the four networks is shown in Fig. 4. We highlight three features. First, our algorithm SelectBP outperforms the other five algorithms in all four networks. It shows that our cross-layer design and backup path selection algorithm are effective for finding reliable backup paths. Not-via ignores the correlation between logical link failures, and thus backup

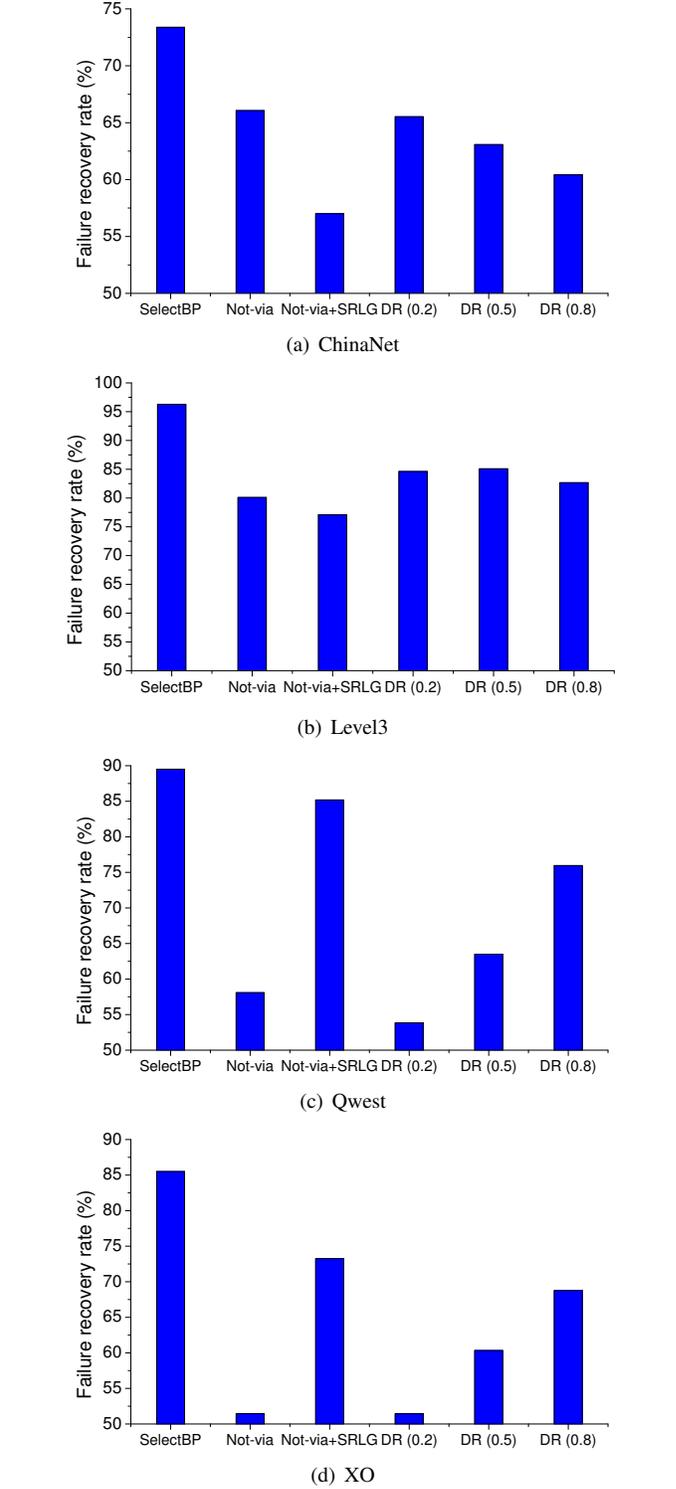(a) ChinaNet

(b) Level3

(c) Qwest

(d) XO

Fig. 3. The average failure recovery rate.

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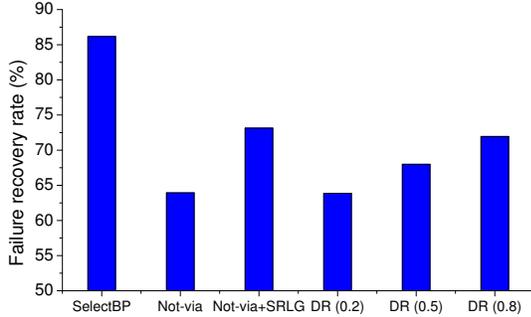

Fig. 4. The overall failure recovery rate for the four networks.

A fiber link failure may cause multiple logical link failures. In some cases, all these failed logical links are recovered; while in some other cases, only some failed logical links are recovered. We count the cases that all failed logical links are recovered. The percentage of these test cases is summarized in Table V. Similar to the failure recovery rate, SelectBP outperforms the other fiver algorithms in all four networks.

TABLE V
PERCENTAGE OF THE TEST CASES THAT FAILED LOGICAL LINKS ARE ALL
RECOVERED

Algorithm	ChinaNet	Level3	Qwest	XO
SelectBP	56.3	92.9	85.2	83.8
Not-via	47.8	62.6	37.0	29.6
Not-via+SRLG	36.8	57.7	75.3	54.0
DR (0.2)	48.8	69.8	35.3	36.6
DR (0.5)	44.4	68.9	40.9	40.5
DR (0.8)	40.8	64.7	55.1	47.7

C. Traffic Recovery Rate and Logical Link Overload

Next, we consider the bandwidth constraint and evaluate the performance of the algorithm SelectBP-BC proposed in Section IV-B. The average logical link utilization of generated traffic is from 5% to 40% in increments of 5%. We use 40% as the upper bound because ISPs usually upgrade their infrastructure when the logical link utilization exceeds 40% [24], [25]. Therefore, it is important to investigate the performance when the logical link utilization is up to 40%. We define the following two metrics for measuring the benefit and negative impact.

- *Traffic recovery rate*: Suppose logical link $e_{m,n}$ fails. If its backup path $B_{m,n}$ does not contain any failed or overloaded logical link, the traffic of $e_{m,n}$ rerouted by $B_{m,n}$ is recovered. The recovered traffic load in SelectBP-BC is $r_{m,n}$, because it controls the rerouted traffic load on $B_{m,n}$. The other five algorithms do not control the rerouted traffic load, and thus the recovered

traffic load is $l_{m,n}$. Suppose the overall traffic load of all failed logical links is T. We measure the recovered traffic load T_r achieved by each algorithm. The traffic recovery rate is defined as $\frac{T_r}{T}$. The optimal value (100%) 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 traffic recovery rate is shown in Fig. 5, and the overall result for the four networks is shown in Fig. 6. We highlight three important aspects of the simulation results. First, SelectBP-BC outperforms the other five algorithms under different logical link utilizations in each network. Second, 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. Third, the traffic recovery rate decreases as the logical link utilization increases due to lack of usable bandwidth. For example, if the logical link utilization is 25%, the usable bandwidth is 3 times the traffic load. However, when it increases to 40%, the usable bandwidth decreases to 1.5 times the traffic load. A small increase in the traffic load makes rerouting much more difficult.
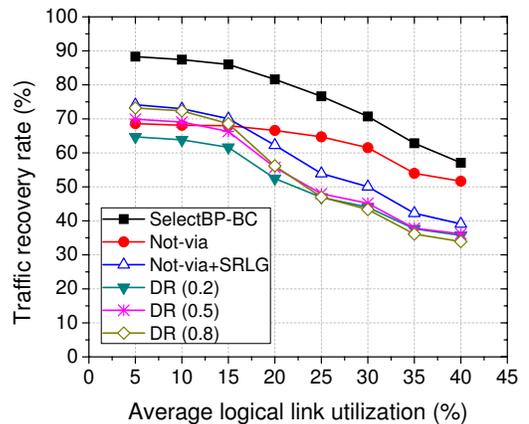

Fig. 6. The overall traffic recovery rate for the four networks.

Next, we evaluate the overload rate under different logical link utilizations. The result in different networks is shown in Fig. 7, and the overall result for the four networks is shown in Fig. 8. SelectBP-BC avoids logical link overload with two techniques, i.e., using logical links with adequate bandwidth and controlling the rerouted traffic load. The other five algorithms may have quite high overload rate when the logical link utilization is above 15%.
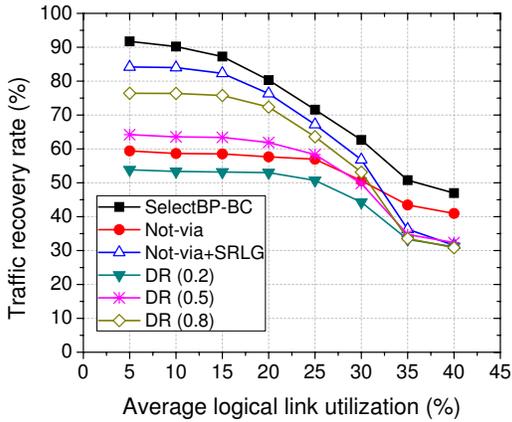
We also measure the maximum logical link utilization on backup paths and show the result when the average logical link
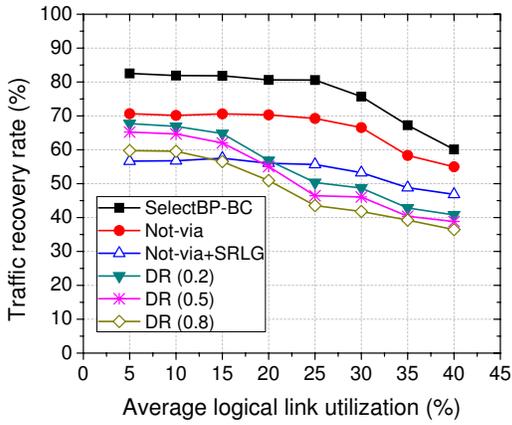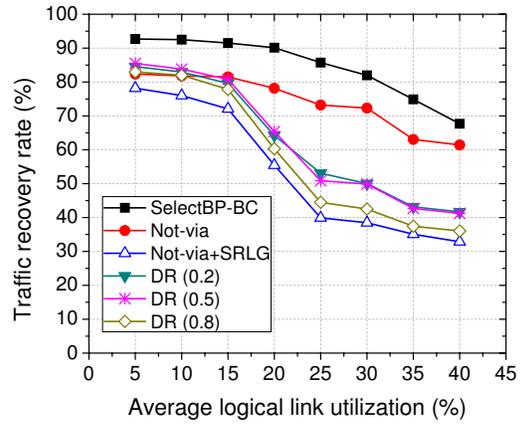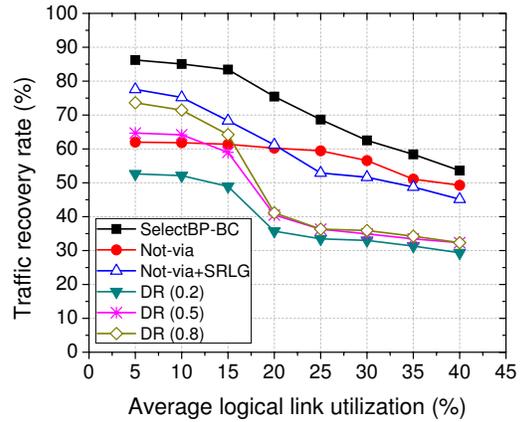
Fig. 5. The average traffic recovery rate under different logical link utilizations..

utilization is 40% in Table VI. By considering the bandwidth constraint, the maximum logical link utilization in SelectBP-BC is 100%; i.e., SelectBP-BC fully utilizes the bandwidth and does not cause logical link overload. The other five algorithms do not consider the bandwidth constraint when rerouting traffic, and then 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, which is even higher than 362%. Improperly rerouting traffic may cause severe congestion and strongly interferes with normal traffic. A survey to ISPs confirms the existence of severe congestion caused by IPFRR in their networks [16].

TABLE VI
THE MAXIMUM LOGICAL LINK UTILIZATION ON BACKUP PATHS WHEN
THE AVERAGE LOGICAL LINK UTILIZATION IS 40%

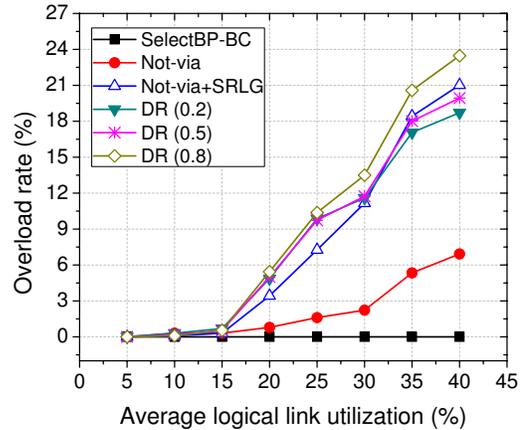| Algorithm	ChinaNet	Level3	Qwest	XO
SelectBP	100	100	100	100
Not-via	165.2	204	341.3	136.4
Not-via+SRLG	228.4	362.8	305.2	252
DR (0.2)	210	279.2	241.3	318.4
DR (0.5)	309	330.4	343.3	275.2
DR (0.8)	212.8	297.2	244.4	260.8

Fig. 8. The overall overload rate for the four networks.

VI. RELATED WORK

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

Backup path-based IP link protection: Most existing works consider the backup path selection as a connectivity problem and mainly focus on finding backup paths to bypass the failed
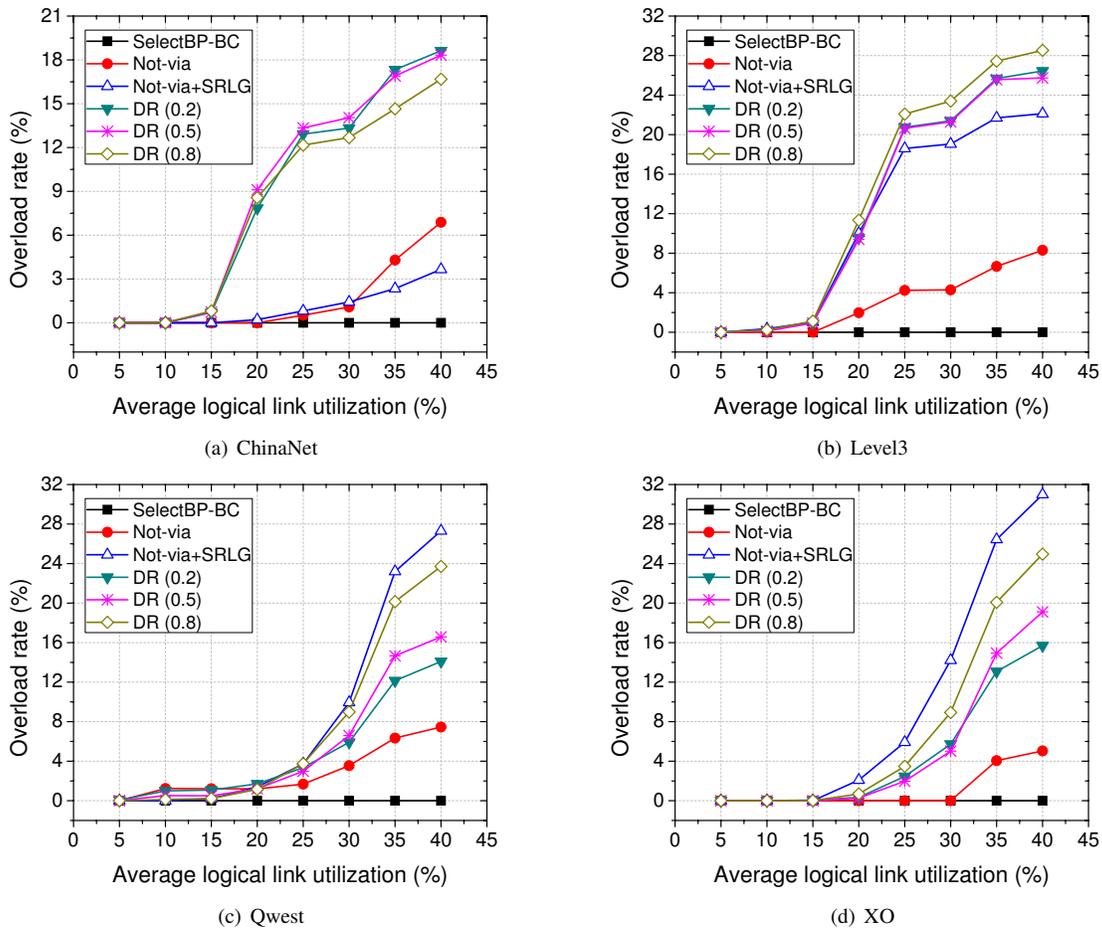
Fig. 7. The overload rate under different logical link utilizations.

IP links [3], [7], [9], [27]. Some other works address the backup path selection when IP links have different failure probability, and aim at maximizing the reliability of backup paths [10], [14]. Zheng *et al.* [37] proposed an approach to on-demand search for the shortest recovery paths to recover intra-domain routing from large-scale failures. However, these approaches ignore the fact that a recovery path may not have enough bandwidth. Consequently, the rerouted traffic may cause severe logical link overload as observed on a major IP backbone by Iyer *et al.* [15]. There are some works on detecting IP link failures [21], [28] and router failures [34], but they do not address the failure recovery.

There are some existing works that consider the traffic distribution among backup paths. Kvalbein *et al.* [29] proposed an approach to distribute the rerouted traffic without causing logical link overload. However, it considers that different backup paths have the same reliability, and it only uses the IP layer information for backup path selection. The approaches in [11], [30] address the logical link overload caused by backup paths. However, they aim at minimizing the bandwidth allocated to backup paths and consider logical link failures as independent events. R3 [16] splits the rerouted traffic among multiple backup paths to minimize the maximum link

utilization. It requires routers to broadcast failure information immediately after detecting a logical link failure. Furthermore, it aims at achieving traffic engineering goals rather than recovery purposes. The approach in [31] jointly addresses the recovery and traffic engineering in multi-path routing. Like R3, it aims at minimizing the maximum link utilization. Moreover, these two solutions do not consider the reliability of the backup paths.

All aforementioned approaches are built on single-layer information, whereas our approach is based on a cross-layer design. Another key difference is that our method jointly considers minimizing traffic disruption due to failures and avoiding logical link overload caused by the rerouted traffic.

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 in IP-over-WDM networks [6], [19], [32], [33], i.e., building the mapping between logical and fiber links to minimize the impact of fiber link failures on logical links. Lee *et al.* [20], [22] analyzed the reliability of IP-over-WDM networks under fiber link failures. They showed that the reliability of IP layer is strongly affected by the correlation between the logical and physical topologies.

However, these works do not address the problem of selecting backup paths to protect IP links. Moreover, they do not model the impact of a logical link failure on the failure probability of fiber links, logical links, and backup paths, which is solved by our CFP model. Chigan *et al.* [35] protected IP routers by reusing optical wavelength channels. It deals with the optical layer configuration but does not consider the backup path selection in the IP layer. Furthermore, it only considers single failure; while our method does not have this assumption. Cui *et al.* [36] 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.

VII. CONCLUSIONS

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 IP link protection. Based on the topology mapping, we develop a correlated failure probability model to quantify the impact of IP link failure on the reliability of backup paths. With this model, we propose a heuristic algorithm to choose the backup paths with minimum failure probability. We also propose a multi-round algorithm that considers the bandwidth constraint in backup path selection. It aims at choosing backup paths to minimize the traffic disruption caused by failures. Moreover, it ensures that the rerouted traffic does not cause logical link overload. We evaluate our approach using real ISP networks with both optical and IP layer topologies. Compared with prior works, the backup paths selected by our method are much more reliable. Moreover, our approach achieves higher recovery rate without causing logical link overload.

REFERENCES

[1] A. Markopoulou, G. Iannaccone, S. Bhattacharyya, C.-N. Chuah, and C. Diot, "Characterization of failures in an IP backbone," in *IEEE INFOCOM*, 2004.

[2] V. Sharma and F. Hellstrand, "Framework for MPLS-based recovery," RFC 3469, 2003.

[3] M. Shand and S. Bryant, "IP fast reroute framework," RFC5714, January 2010.

[4] P. Francois, C. Filsfils, J. Evans, and O. Bonaventure, "Achieving sub-second IGP convergence in large IP networks," *ACM SIGCOMM Computer Communication Review*, vol. 35, no. 3, 2005.

[5] F. Giroire, A. Nucci, N.Taft, and C. Diot, "Increasing the robustness of IP backbones in the absence of optical level protection," in *IEEE INFOCOM*, 2003.

[6] O. Crochat and J.-Y. L. Boudec, "Design protection for WDM optical networks," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 7, 1998.

[7] A. Kvalbein, A. F. Hansen, T. Cicic, S. Gjessing, and O. Lysne, "Fast IP network recovery using multiple routing configurations," in *IEEE INFOCOM*, 2006.

[8] 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*, vol. 15, no. 2, 2007.

[9] S. Kini, S. Ramasubramanian, A. Kvalbein, and A. F. Hansen, "Fast recovery from dual link failures in IP networks," in *IEEE INFOCOM*, 2009.

[10] M. Hou, D. Wang, M. Xu, and J. Yang, "Selective protection: A cost-efficient backup scheme for link state routing," in *IEEE ICDCS*, 2009.

[11] M. Johnston, H.-W. Lee, and E. Modiano, "A robust optimization approach to backup network design with random failures," in *IEEE INFOCOM*, 2011.

[12] E. Oki, N. Matsuura, K. Shiomoto, and N. Yamanaka, "A disjoint path selection scheme with shared risk link groups in GMPLS networks," *IEEE Communications Letters*, vol. 6, no. 9, 2002.

[13] 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," *IEEE/ACM Transactions on Networking*, vol. 13, no. 4, 2005.

[14] H.-W. Lee and E. Modiano, "Diverse routing in networks with probabilistic failures," in *IEEE INFOCOM*, 2009.

[15] S. Iyer, S. Bhattacharyya, N. Taft, and C. Diot, "An approach to alleviate link overload as observed on an IP backbone," in *IEEE INFOCOM*, 2003.

[16] Y. Wang, H. Wang, A. Mahimkar, R. Alimi, Y. Zhang, L. Qiu, and Y. R. Yang, "R3: Resilient routing reconfiguration," in *ACM SIGCOMM*, 2010.

[17] A. Basu and J. Riecke, "Stability issues in OSPF routing," in *ACM SIGCOMM*, 2001.

[18] E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol label switching architecture," RFC 3031, January 2001.

[19] E. Modiano and A. Narula-Tam, "Survivable lightpath routing: A new approach to the design of WDM-based networks," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 4, 2002.

[20] K. Lee and E. Modiano, "Cross-layer survivability in WDM-based networks," in *IEEE INFOCOM*, 2009.

[21] Q. Zheng and G. Cao, "Minimizing probing cost and achieving identifiability in probe based network link monitoring," *IEEE Transactions on Computers*, to appear.

[22] K. Lee, H.-W. Lee, and E. Modiano, "Reliability in layered networks with random link failures," in *IEEE INFOCOM*, 2010.

[23] 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 Transactions on Networking*, vol. 16, no. 4, 2008.

[24] S. Kandula, D. Katabi, B. Davie, and A. Charny, "Walking the tightrope: Responsive yet stable traffic engineering," in *ACM SIGCOMM*, 2005.

[25] J. Guichard, F. le Faucheur, and J. P. Vasseur, *Definitive MPLS network design*. Cisco Press, 2005.

[26] B. Fortz and M. Thorup, "Internet traffic engineering by optimizing ospf weights," in *IEEE INFOCOM*, 2000.

[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] Q. Zheng and G. Cao, "Minimizing probing cost and achieving identifiability in network link monitoring," in *IEEE ICDCS*, 2010.

[29] A. Kvalbein, T. Cicic, and S. Gjessing, "Post-failure routing performance with multiple routing configurations," in *IEEE INFOCOM*, 2007.

[30] R. Banner and A. Orda, "Designing low-capacity backup networks for fast restoration," in *IEEE INFOCOM*, 2010.

[31] M. Suchara, D. Xu, R. Doverspike, D. Johnson, and J. Rexford, "Network architecture for joint failure recovery and traffic engineering," in *ACM SIGMETRICS*, 2011.

[32] M. Kurant and P. Thiran, "On survivable routing of mesh topologies in IP-over-WDM networks," in *IEEE INFOCOM*, 2005.

[33] 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 *IEEE INFOCOM*, 2009.

[34] Q. Zheng, G. Cao, T. L. Porta, and A. Swami, "Detecting and localizing large-scale router failures using active probes," in *IEEE MILCOM*, 2011.

[35] C. Chigan, G. W. Atkinson, and R. Nagarajan, "Cost effectiveness of joint multilayer protection in packet-over-optical networks," *Journal of Lightwave Technology*, vol. 21, no. 11, 2003.

[36] W. Cui, I. Stoica, and R. H. Katz, "Backup path allocation based on a correlated link failure probability model in overlay networks," in *IEEE ICNP*, 2002.

[37] Q. Zheng, G. Cao, T. L. Porta, and A. Swami, "Optimal recovery from large-scale failures in IP networks," in *IEEE ICDCS*, 2012.