Copyright ⓒ 2014 KSII

Load-Balanced One-hop Overlay Multipath Routing with Path Diversity

Jianxin Liao[1*], Shengwen Tian[1, 2], Jingyu Wang[1], Tonghong Li[3] and Qi Qi[1]
[1] State Key Laboratory of Networking and Switching Technology,
Beijing University of Posts and Telecommunications, Beijing, 100876 - China
[e-mail: {liaojx, wangjingyu, qiqi8266}@bupt.edu.cn]
[2] School of Information and Electrical Engineering, Ludong University
Yantai, 264000 - China
[e-mail: shwtian@gmail.com]
[3] Department of Computer Science, Technical University of Madrid
Madrid, 28000 - Spain
[e-mail: tonghong@fi.upm.es]
*Corresponding author: Jianxin Liao

Received October 15, 2013; revised January 4, 2014; accepted February 9, 2014; published February 28, 2014

Abstract

Overlay routing has emerged as a promising approach to improve reliability and efficiency of the Internet. For one-hop overlay source routing, when a given primary path suffers from the link failure or performance degradation, the source can reroute the traffic to the destination via a strategically placed relay node. However, the over-heavy traffic passing through the same relay node may cause frequent package loss and delay jitter, which can degrade the throughput and utilization of the network. To overcome this problem, we propose a Load-Balanced One-hop Overlay Multipath Routing algorithm (LB-OOMR), in which the traffic is first split at the source edge nodes and then transmitted along multiple one-hop overlay paths. In order to determine an optimal split ratio for the traffic, we formulate the problem as a linear programming (LP) formulation, whose goal is to minimize the worse-case network congestion ratio. Since it is difficult to solve this LP problem in practical time, a heuristic algorithm is introduced to select the relay nodes for constructing the disjoint one-hop overlay paths, which greatly reduces the computational complexity of the LP algorithm. Simulations based on a real ISP network and a synthetic Internet topology show that our proposed algorithm can reduce the network congestion ratio dramatically, and achieve high-quality overlay routing service.

Keywords: One-hop overlay routing, Load balancing, Linear programming, Betweenness centrality, Multipath, Path diversity

This work was jointly supported by: (1) the National Basic Research Program of China (No. 2013CB329102); (2) National Natural Science Foundation of China (No. 61372120, 61271019, 61101119, 61121001, 61072057, 60902051); (3) PCSIRT (No. IRT1049); (4) Beijing Higher Education Young Elite Teacher Project (Grant Nos. YETP0473); (5) MICINN (No.TIN2010-19077); (6) CAM (No.S2009TIC-1692).

http://dx.doi.org/10.3837/tiis.2014.02.007

1. Introduction

Link and router failures are frequent in the Internet [1][2]. The convergence time for routing protocols to route around these failures is often in the order of seconds or minutes [3][4], during which certain end-to-end connections may experience seconds or minutes of outage [5]. Overlay routing has been proposed in recent years as an effective way to improve reliability and efficiency of the Internet without any changes in the Internet infrastructure. For example, overlay routing has been used to improve the reliability of Internet paths in RON (Relisient Overlay Network)[6][7]. It has also been used for providing Internet QoS in QRON (QoS-aware Routing in Overlay Networks) [8]. In overlay routing, an end host has the flexibility in routing its traffic to its destination through one or multiple overlay relay nodes.

When a given physical path suffers from the link failure or performance degradation, the source can reroute the traffic to the destination relayed by an overlay node to detour the failed links, which is called one-hop overlay source routing [9]. In one-hop overlay source routing, an overlay path consists of two overlay links, and each overlay link consists of one or multiple physical links, as shown in **Fig. 1**.

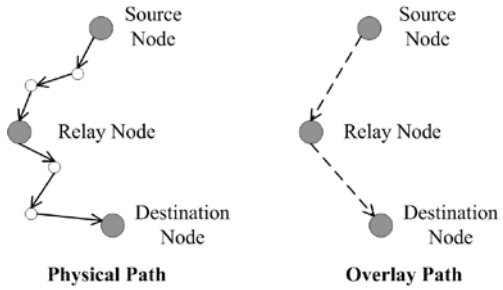


Fig. 1. One-hop overlay source routing scheme

The problem of one-hop overlay source routing has been discussed in previous literatures [9], [10], and [11]. These researches concentrate on single-path overlay routing without considering load balancing. In other words, the traffic between each source-destination node pair is relayed by one intermediate overlay node. However, with the rapid increase of new Internet-based applications, such as voice-over-IP, peer-to-peer, and video-on-demand, large amount of multimedia data need to be transmitted between source-destination node pairs. In such a case, multipath transmission can increase the throughput of network. On the other hand, if the traffic between different source-destination pairs passes through the same intermediate overlay node simultaneously, it may cause frequent package loss and delay jitter, which can degrade the throughput and utilization of the network.

This paper is the first devoted to overcome this problem, and proposes a Load-Balanced One-hop Overlay Multipath Routing algorithm with path diversity (LB-OOMR). In LB-OOMR, when a path failure is detected, the source node selects k ($k \geq 2$) overlay relay nodes to construct k one-hop overlay alternative paths, and then split its traffic into k sub-traffics, and reroute these sub-traffics through the constructed k one-hop overlay paths. Note that the traffic is rerouted from the source to the relay nodes and from the relay nodes to the destination along the shortest path.

The selection of one-hop overlay routing paths has a drastic effect on the performance of LB-OOMR. To increase the reliability and robustness of the network, it is desirable and

Copyright ⓒ 2014 KSII

beneficial to take advantage of path diversity to select one-hop overlay routing paths, which minimizes the number of joint physical links among k +1 routing paths including one default physical path and k one-hop overlay paths. Therefore, even if a sub-path fails, the traffic is still able to reach the destination through other paths, which guarantees the robustness of the network. In LB-OOMR, we not only take advantage of path diversity to select the overlay relay nodes for establishing k $(k \geq 2)$ one-hop overlay paths, but also consider the capacity of node and link for load balancing.

The key to load balancing is how to allocate the traffic over each one-hop overlay path, i.e., to determine an optimal split ratio. To solve this problem, a linear programming (LP) formulation is developed, whose goal is to minimize the worse-case network congestion ratio. Since it is difficult to solve this LP problem in practical time, a heuristic algorithm is proposed. Because the selection of overlay relay nodes can influence directly the complexity and performance of the LP optimization algorithm, our heuristic algorithm concentrates on this issue.

For the selection of overlay relay nodes, spurred by the characteristics that a few nodes with high betweenness centrality can provide more optimal routes for a large number of node pairs in the Internet [10][12], we select a given number of overlay nodes whose betweenness centralities are higher than others as the candidate overlay relay nodes. k $(k \geq 2)$ overlay relay nodes are selected from the candidate relay nodes to construct k one-hop overlay paths, which is beneficial to reduce the search space and improve the performance of LP optimization algorithm.

The rest of the paper is organized as follows. In Section II, we introduce the related work. Section III presents the network model and the terminologies used in this paper. Our proposed LB-OOMR algorithm is described in Section IV. In Section V, we present the simulation results and analyze the performance of LB-OOMR. Finally, the paper is concluded in Section VI.

2. Related Work

There have been considerable researches on overlay routing to improve the reliability and performance of the Internet. Reference [13] shows that in 30%-80% of the Internet routing paths there is an alternate routing path with better quality compared to the default routing path. RON [6] is a one-hop overlay routing method, which quickly detects and recovers path outages and the degraded performance. But RON lacks the scalability and does not consider load balancing. In [14], the authors study the one-hop overlay routing problem for the robustness of the network, but only focus on the placement of relay nodes in an intra-domain network. In SOSR (Scalable One-hop Source Routing) [9], the authors present the concept of one-hop source routing and study this problem by using the experiment data on the PlanetLab. The results in SOSR show that one-hop source routing with four relay nodes selected randomly from the network can recover from 56% of network failures. References [10] and [11] study the cost associated with the relay node placement for overlay routing. In addition, in our earlier work [15], we proposed an open multi-plane framework for Next Generation Service Overlay Network (NGSON), in which different functional overlays can systematically be coordinated with each other.

Many researches on load-balanced routing have been conducted. Multipath routing schemes with load balancing can be classified into traditional IP-based and multiprotocol label switching (MPLS) based. The IP-based multipath routing needs to extend the existing routing

algorithms (RIP, OSPF, or BGP) for multipath support, which cannot take full advantage of multiple paths that frequently exist in Internet Service Provider Network [16]. Although the MPLS-based multipath routing [17][18] is proposed as a powerful technology supporting load balancing recently, the sophisticated operations are performed by the Multi-Protocol Label Switching (MPLS) Traffic-Engineering (TE) technology, which focuses on the IP-layer network. However, legacy networks mainly employ shortest-path-based routing protocols such as Open Shortest Path First (OSPF) and Intermediate System to Intermediate System (IS-IS). This means that the IP routers deployed in the legacy networks need to be transformed into Label Switching Routers (LSR) for supporting Label Distribution Protocol (LDP), which will significantly increase the capital expenditures [19]. In addition, TE needs to change the routing path frequently to adapt the dynamic traffic demand, which may cause the network instability. Different from the previous literatures, our proposal is deployed at the application layer without any changes in the Internet infrastructure.

3. Network Model

In this paper, the physical network is represented as a directed graph $G(V,E)$, where V is the set of nodes and E is the set of links. The sets of incoming and outgoing edges at node i are denoted by $E^-(i)$ and $E^+(i)$, respectively. Let $(i,j)\in E$ represent a directed link in the network from node $i\in V$ to node $j\in V$. To simplify the notation, we also refer to a link by e instead of (i,j). C_{ij} and L_{ij} is the capacity and load of link (i,j), respectively. The overlay nodes are given as a subset $Q\subseteq V$ where each node can be a source or destination of traffic. Let $|Q|=N$. For each $i\in Q$, we denote the upper bounds on the total amount of traffic entering and leaving node i by $b^-(i)$ and $b^+(i)$ respectively, which can avoid overload on the node i. Let d_{ij} represent the traffic between nodes i and j. Any allowable traffic matrix $T=(d_{ij})_{i,j\in Q}$ for the network must obey:

$$\sum_{j\in Q} d_{ij} \le b^+(i) \quad and \quad \sum_{j\in Q} d_{ji} \le b^-(i) \tag{1}$$

We assume that $d_{ii}=0$ for all nodes $i\in Q$.

The network congestion ratio μ refers to the maximum value of all link utilization rates in the physical network. μ is defined by,

$$\mu = \max_{(i,j)\in E}\left\{\frac{L_{ij}}{C_{ij}}\right\} \tag{2}$$

where $0\le\mu\le 1$. Minimizing μ means that the admissible traffic is maximized. Thus, minimizing μ through routing control is the objective of this paper. The notations used in this paper are summarized in **Table 1**.

Copyright ⓒ 2014 KSII

Table 1. Notations

Notations	Description
$e=(i, j)$	Physical link.
Q	Set of overlay nodes, $Q \subseteq V$.
C_{ij}	Capacity of link (i, j) .
L_{ij}	Load of link (i, j) .
d_{ij}	Traffic demand on node pair (i, j) .
$E^{-}(i)$, $E^{+}(i)$	Set of incoming and outgoing edges at node i .
$b^{+}(i)$, $b^{-}(i)$	Bounds of traffic that node i can send into and receive from the network.
δ_m^{pq}	Fraction of traffic demand d_{pq} relayed by the intermediate node $m \in Q$.
ψ_{pq}^{ij}	Link indictor to indict whether the shortest path from node p to node q includes link (i, j) .
$BC(v)$	Betweenness centrality of node v .
σ_{st}	Number of shortest paths from s to t
$\sigma_{st}(v)$	Number of shortest paths from s to t that go through v .
N	Number of overlay nodes.
M	Number of candidate relay nodes, $M \le N$.
I	Set of candidate relay nodes, $I \subseteq Q$.
k	Number of relay nodes for one-hop overlay routing, $k \le M$.
R	Set of relay nodes for one-hop overlay routing, $R \subseteq I$.
μ	Congestion ratio.
$\mu_{LB-OOMR}$	Congestion ratio obtained by our proposed algorithm LB-OOMR.
μ_{RSM}	Congestion ratio obtained by Random Selection Method (RSM).
μ_{SND}	Congestion ratio obtained by Selection method based on Node Degree (SND).
μ_{non}	Congestion ratio obtained by non-split one-hop overlay routing.

4. Load-Balanced One-hop Overlay Multipath Routing with Path Diversity

LB-OOMR is conceptually straightforward. When a path failure is detected, the source node first selects k ($k \ge 2$) overlay relay nodes to construct k one-hop overlay alternative paths, and then split its traffic into k sub-traffics, and reroute these sub-traffics through the constructed k different one-hop overlay paths. During the rerouting, each sub-traffic is transferred between a source-destination node pair in two stages. First, k sub-traffics are directed to k different overlay relay nodes, respectively. Next, every relay node forwards the received sub-traffic to the final destination. The traffic is first routed from the source to the relay nodes and then from the relay nodes to the destination according to the shortest-path-based protocol in the physical network. For example in **Fig. 2**, when the source p suffers from a path failure to the destination q , its traffic is split into four sub-traffics (i.e. $k=4$) and rerouted simultaneously through relay nodes m_1 , m_2 , m_3 , m_4 .

LB-OOMR determines the split ratio for each source-destination pair independently. To determine a set of optimal split ratio that minimizes the network congestion ratio, a general LP formulation is presented as follows.

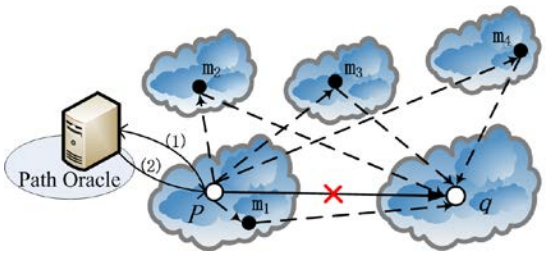


Fig. 2. One-hop overlay source routing with multiple paths

4.1 LP Formulation

Different source-destination pair has a different set of k relay nodes. For each traffic demand d_{pq} routed from source node $p \in Q$ to destination node $q \in Q$, we define δ_m^{pq} as the fraction of traffic from p to q relayed by the relay node $m \in Q$ in the one-hop overlay network.

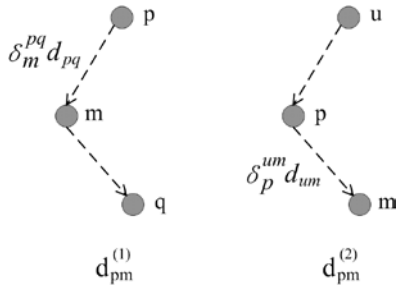


Fig. 3. Traffic distribution of LB-OOMR

We assume that d_{pm} refers to the traffic between node p and node m, which consists of two components, as shown in **Fig. 3**. The first one is the traffic generated by node p and relayed by node m, which is defined as $d_{pm}^{(1)}$. The second one is the traffic for m relayed by node p, which is defined as $d_{pm}^{(2)}$. In other words, node p is the source and node m is the relay node in $d_{pm}^{(1)}$. While in $d_{pm}^{(2)}$ node p is the relay node and node m is the destination. It is easy to see that $d_{pm}^{(1)}$ and $d_{pm}^{(2)}$ hold:

$$d_{pm}^{(1)} = \sum_{q \in Q} \delta_m^{pq} d_{pq} \tag{3}$$

$$d_{pm}^{(2)} = \sum_{u \in Q} \delta_p^{um} d_{um} \tag{4}$$

Therefore, d_{pm} is given by:

$$d_{pm} = d_{pm}^{(1)} + d_{pm}^{(2)} = \sum_{q \in Q} \delta_m^{pq} d_{pq} + \sum_{u \in Q} \delta_p^{um} d_{um} \tag{5}$$

In the same way, d_{mq} is represented as follows:

Copyright ⓒ 2014 KSII

$$d_{mq} = d_{mq}^{(1)} + d_{mq}^{(2)} = \sum_{p \in Q} \delta_m^{pq} d_{pq} + \sum_{v \in Q} \delta_q^{mv} d_{mv} \tag{6}$$

Let $\psi_{pm}^{ij} = 1$ if the shortest path from node p to node m traverses through the link (i, j), and $\psi_{pm}^{ij} = 0$ otherwise.

Now we are ready to present our algorithm. For each source-destination pair (p,q), we need to determine k one-hop overlay routing paths and the optimal split ratio δ_m^{pq} on each one-hop overlay path. Let $\bigcup(m)$ be the total number of the selected relay node m, that is $\bigcup(m) = k$. The main idea is to formulate the problem as a linear programming, which can be stated as follows:

minimize: μ (7)

s.t.

$$\sum_{m \in Q} \delta_m^{pq} = 1 \tag{8}$$

$$\sum_{m \in Q} (\psi_{pm}^{ij} + \psi_{mq}^{ij}) \delta_m^{pq} d_{pq} + \beta_{ij} \le \mu C_{ij}, \quad p \ne q \ne m \tag{9}$$

$$\sum_{(u,j) \in E^+(u)} \psi_{pq}^{uj} - \sum_{(i,u) \in E^-(u)} \psi_{pq}^{iu} = \begin{cases} +1, & if \quad u = p \\ -1, & if \quad u = q \\ 0, & otherwise \end{cases} \tag{10}$$

$$\sum_{q \in Q} d_{mq} \le b^+(m), \quad m \in Q \tag{11}$$

$$\sum_{p \in Q} d_{pm} \le b^-(m), \quad m \in Q \tag{12}$$

$$0 \le \delta_m^{pq} \le 1 \tag{13}$$

$$0 \le \mu \le 1 \tag{14}$$

$$\bigcup(m) = k \tag{15}$$

In LP (7)-(15), p, q and m denotes the source node, the destination node and the relay node. The objective function in Eq. (7) minimizes the network congestion ratio, i.e. maximizes the throughput of the network. Constraint (8) states that the sum of δ_m^{pq} through all relay nodes m for each source-destination node pair in the one-hop overlay network is equal to 1. Constraint (9) requires that the utilization of each physical link on one-hop overlay path cannot exceed the congestion ratio μ. $\psi_{pm}^{ij} \in \{0,1\}$ and $\psi_{mq}^{ij} \in \{0,1\}$. When $\psi_{pm}^{ij} = 1$ and $\psi_{mq}^{ij} = 1$, the physical link (i, j) simultaneously belongs to the overlay link (p,m) and (m,q). In constraint (9), β_{ij} is the background traffic of the link (i, j), which can be obtained from the

traffic matrix. The values d_{pq} and C_{ij} in constraint (9) are constants, and hence this constraint is linear. Constraint (10) is the flow conservation constraint, ensuring that the variable ψ_{pq}^{ij} represents a flow of value 1 from p to q in the overlay network. Constraint (11) and (12) are the limitation of out- and in-traffic of the relay nodes in the overlay network, in which d_{pm} and d_{mq} depend on Eq. (5) and (6), respectively. Constraint (13) and (14) give the bounds for the variables. Constraint (15) requires that the number of relay nodes is k, i.e., the number of one-hop overlay paths is k.

Because we need to select simultaneously k relay nodes for one-hop overlay multipath routing, the selection process of these k relay nodes is not independent. So, the time complexity of LP (7)-(15) is equivalent to the combination number C_N^k. With the increase of the number of overlay nodes N, it becomes harder to solve the LP problem within a practical time. Therefore, a heuristic algorithm is required.

4.2 Heuristic Algorithm

In LB-OOMR, as the traffic is routed from the source to the relay nodes and from the relay nodes to the destination along the shortest path, for each source-destination pair, if the traffic to a destination is routed via a predefined set of relay node, the time complexity of LP (7)-(15) is reduced to $O(1)$. In the pursuit of this endeavor, we divide LB-OOMR into two steps. In the first step, we select k suitable relay nodes from the set Q for constructing k one-hop overlay routing paths. In the second step, we compute the fraction of traffic δ_m^{pq} on each sub-path for minimizing the congestion ratio μ. We introduce a heuristic algorithm to concentrate on the first step, in which we first define a set of candidate relay nodes, and then select strategically k relay nodes from the set of candidate nodes.

4.2.1 Selection of candidate relay nodes

In order to reduce the search space of the LP (7)-(15), we first define a set of candidate relay nodes $I \subseteq Q$ for the selection of relay nodes, as shown in **Fig. 4**.

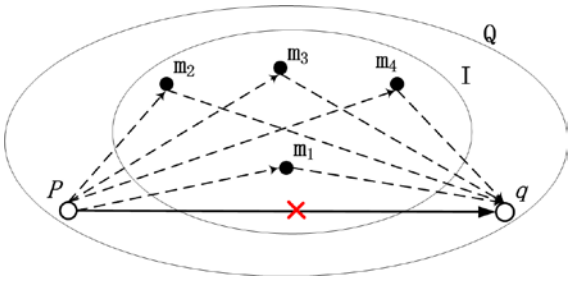


Fig. 4. Selection of candidate relay nodes

Our ideas come from the characteristics, in which only a few nodes with high betweenness centrality are repeatedly present in many routing paths [10]. In other words, a small number of relay nodes can provide optimal routes to a large portion of end-to-end pairs. Betweenness centrality [20] of a node v is the sum of the fraction of all-pairs shortest paths that pass through v, which is denoted as follows:

Copyright ⓒ 2014 KSII

$$BC(v) = \sum_{s,t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{16}$$

where V is the set of nodes, σ_{st} denotes the number of shortest paths from s to t, and for any $v \in V$, $\sigma_{st}(v)$ is the number of shortest paths from s to t that go through v.

In order to validate this characteristics, we use the data of a real Internet topology CN070 [21] (depicted in detail in Section V) and plot the betweenness centralities of all nodes in the network, as shown in **Fig. 5**, where in x-axis node IDs are sorted by their betweenness centralities in a decreasing order. In CN070, the link bandwidth (available bandwidth) is assigned according to a uniform distribution in the range [40, 120] Mb/s. Assigning different weights to the links can generate different network topologies. In **Fig. 5**, the betweenness centrality of each node is the average value after assigning the link weight for 2000 times. From this figure, we can obtain that only a few nodes have extremely high betweenness centralities.

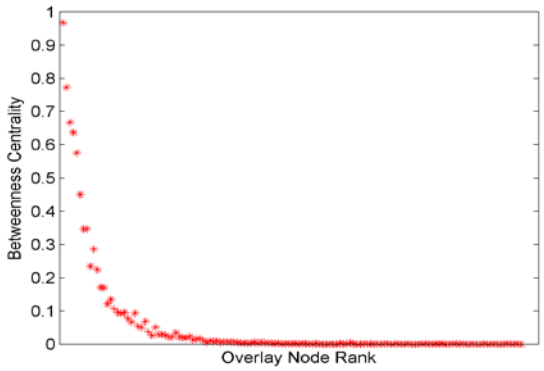


Fig. 5. Betweenness centralities of nodes in the physical network

We select the nodes with higher betweenness centrality as the candidate relay nodes, which can reduce the hops of routing path between each source-destination node pair. To some extent, smaller routing hops means shorter latency. Therefore, we compute the betweenness centrality of each overlay node, and select M nodes with highest betweenness centralities in the overlay network as the candidate relay nodes and form the set I. The size of M depends on the size of physical network; the experiment data (in Section V) show that about 10% of the network size can achieve a good effect. When the arrival or departure of some overlay nodes causes the changes of the set Q, we recalculate the betweenness centrality of each overlay node in Q to update the set I.

4.2.2 Selection of k relay nodes

The problem of selecting k relay nodes is to find a set $R \subseteq I$ of k nodes such that the overlap among the k +1 routing paths including the default physical path and k one-hop overlay paths is minimized. The overlap between two paths is defined as the number of joint physical links that are common between these two paths. Reference [22] shows that the failure correlation of two paths depends on the extent of how much they overlap.

Let S be a set of k nodes selected from the set I. Thus we can obtain C_M^k different S, in which M denotes the size of the set I. For each S, we define $p(S)$ is the set of k one-hop overlay paths relayed by k different intermediate node $m_i \in S$, $i = 1,2,3,\cdots,k$. $p(S)$ can be

represented by $p(S)=\bigcup_{i=1}^{k} p(u,m_i,v)$, where $p(u,m_i,v)$ denotes the one-hop overlay path from the source u to the destination v relayed by the intermediate node $m_i \in S$. Therefore, the set R can be represented as the following equation:

$$R = arg\min_{S} LO[p(S), p^*(u,v)] \tag{17}$$

where $p^*(u,v)$ denotes the default Internet path between u to v, and $LO[p(S), p^*(u,v)]$ is the average pairwise overlap between the set of $k+1$ paths, namely, one direct physical path from u to v, and k one-hop overlay paths between the same two nodes.

While it is important to find one-hop overlay paths with minimum overlaps with the default physical path, it is also imperative that the one-hop overlay paths themselves have as low pairwise overlaps among themselves as possible. The factor $LO[p(S), p^*(u,v)]$ in Eq. (17) is able to capture this reliability feature, by which all the one-hop overlay paths have the minimum failure correlation and provide the reliability under multiple failures. In the meanwhile, the value of k is critical. It should be not too small; otherwise, it is not good for load balancing. And it should not be too large because it is impractical and inefficient to detour data through such a large number of alternative paths. A suitable choice for the value of k is 4, as shown in [9] based on Internet experiments.

We apply an incremental heuristic method to compute the set R, in which one new relay node is selected from the set I at each step. The choice of such a relay node is based on minimizing the objective function $LO[p(S), p^*(u,v)]$. The steps of the method are as follows. If n $(n<k)$ nodes have already been selected from the candidate set I as relay nodes, i.e., $|R|=n$, to select the $(n+1)$-th relay node, we iterate over the remaining $M-n$ candidate nodes. At each iteration, we add one node to the set R and recalculate the objective function $LO[p(S), p^*(u,v)]$ for the new set R. The node that gives the minimum value of the objective function $LO[p(S), p^*(u,v)]$ is chosen as the $(n+1)$-th relay node. We repeat the above process until $|R|=k$.

The complexity of the incremental heuristic method is $O(k \cdot M \cdot d)$, where d is the diameter of the network. $O(d)$ comes from the calculation of overlap that involves finding the set of common links between the default physical path and the one-hop overlay detoured path. For increasing one relay node into the set R, $O(M)$ comes from the calculation of the objective function $LO[p(S), p^*(u,v)]$ by $M-n$ times, i.e., once for each potential relay node. The above steps have to be done k times, where each time it is done, one relay node is selected to be part of the set R.

According to the analysis above, the algorithm of the selection of relay nodes can be described as follows:

Algorithm 1 Selection of Relay Nodes

Input: $G(V,E)$, Q, source-destination pair (p,q) and d_{pq}.

Output: a relay nodes set R.

1: compute the betweenness centralities *BCs* of all overlay nodes in $Q-\{p,q\}$ based on

Copyright ⓒ 2014 KSII

Eq. (16).
2: select M nodes as the candidate relay nodes according to the descending order of *BCs*, and obtain the candidate relay nodes set I.
3: for each node $v \in I$, compute the number of overlap and obtain R based on Eq. (17).

4.2.3 Computing the fraction of traffic δ_m^{pq}

Different source-destination node pair has different relay nodes set R for one-hop overlay routing. After determining the relay nodes set R, the linear programming (7)-(15) can be converted as the following form:

$$\text{minimize:} \quad \mu \tag{18}$$

s.t.

$$\sum_{m \in R} \delta_m^{pq} = 1 \tag{19}$$

$$\sum_{m \in R} (\psi_{pm}^{ij} + \psi_{mq}^{ij}) \delta_m^{pq} d_{pq} + \beta_{ij} \le \mu C_{ij}, \quad p \ne q \tag{20}$$

$$\sum_{(u,j) \in E^+(u)} \psi_{pq}^{uj} - \sum_{(i,u) \in E^-(u)} \psi_{pq}^{iu} = \begin{cases} +1, & if \quad u = p \\ -1, & if \quad u = q \\ 0, & otherwise \end{cases} \tag{21}$$

$$\sum_{q \in Q} d_{mq} \le b^+(m), \quad m \in R \tag{22}$$

$$\sum_{p \in Q} d_{pm} \le b^-(m), \quad m \in R \tag{23}$$

$$0 \le \delta_m^{pq} \le 1 \tag{24}$$

$$0 \le \mu \le 1 \tag{25}$$

Compared to LP (7)-(15), for each traffic demand d_{pq} in LP (18)-(25), k relay nodes have been determined based on the method of the selection of relay nodes, i.e., $m \in R$ instead of $m \in Q$, as shown in constraints (19), (20), (22) and (23), which greatly reduces the computational complexity of the LP algorithm. We just need to compute the split coefficient δ_m^{pq} according to LP (18)-(25), which can be solved optimally with a standard LP solver.

4.3 Deployment of LB-OOMR

For the deployment of LB-OOMR, it is essential to obtain some information about the physical network, such as the network topology and the traffic matrix. To obtain the information, we need to deploy an entity (Path Oracle) in the physical network, as shown in **Fig. 2**. The implementation of Path Oracle can refer to the previous literatures [23][24][25][26]. The Path Oracle acts as an abstract routing underlay to the overlay network, which is a service offered by the ISPs. The oracle service can be realized as a set of replicated servers within each ISP, that is, we might deploy a server in each AS to collect some

information about the AS topology and the network performance. So, the Path Oracle is implemented in a distributed and asynchronous manner.

When the source p detects a path failure to the destination q, the source first sends the request to the Path Oracle with the parameters, including the destination node and the traffic demand, and requests the Path Oracle to provide it with the addresses of k relay nodes and the corresponding split coefficient δ_m^{pq}, cf. Step 1 in **Fig. 2**. Next, Path Oracle obtains the results calculated by LB-OOMR algorithm and returns them to the requester, cf. Step 2 in **Fig. 2**. Finally, the source p uses the received results to forward the traffic to the destination q via k relay nodes.

5. Performance Evaluation

5.1 Simulation Settings

To evaluate the performance of our proposed algorithm LB-OOMR, we compare it with two methods: Random Selection Method (RSM) and Selection based on Node Degree (SND). RSM and SND are designed just for the selection of k relay nodes from the set of overlay nodes Q. The corresponding congestion ratio μ_{RSM}, μ_{SND} and the split coefficient δ_m^{pq} are computed based on LP (18)-(25). The RSM algorithm selects k relay nodes randomly from the set Q, while the SND algorithm greedily chooses k nodes with larger numbers of edges attached to them as the relay nodes from the set Q. Note that the SND algorithm uses the degree of nodes based on the routing edges in the physical network. In addition, we also compute the non-split one-hop overlay routing and obtain its congestion ratio μ_{non}, in which the number of relay node is 1 and the relay node is selected randomly from the overlay nodes. Since the optimal (minimum) congestion ratio μ implies the maximum admissible network traffic, we define $S = 1/\mu$, that is, $S_{LB-OOMR} = 1/\mu_{LB-OOMR}$, $S_{RSM} = 1/\mu_{RSM}$, $S_{SND} = 1/\mu_{SND}$ and $S_{non} = 1/\mu_{non}$.

We carry out the simulations on top of two IP-layer topologies: a real topology CN070 [21] with 135 nodes and 338 links, and a random topology GT180 generated by GT-ITM [27] with 200 nodes and 502 links. CN070 records the interconnection situation of most routers in China in 2006. GT180 is based on the Waxman probability [28]: $P(u,v) = \alpha e^{-d(u,v)/\beta L}$. In the simulation, we take $\alpha = \beta = 0.03$, $L = \sqrt{2}a$ and $a = 180$.

In CN070 and GT180, link capacities are generated randomly with uniform distribution in the range of [80,120]. d_{pq} is also generated randomly with uniform distribution in the range of [0,100]. $b^+(m)$ and $b^-(m)$, which are the capacities of overlay nodes, are also randomly generated in the range of [100, 200]. We set $b^+(m) = b^-(m)$ for each overlay relay node. The link weights used for shortest path computation and betweenness centralities computation are set to be $1/(C_{ij} - L_{ij})$. We set the number of relay nodes k =4, and select randomly a certain number of nodes from the physical network CN070 and GT180 as the set of overlay nodes Q, respectively. In each simulation, we randomly choose a pair of source and destination from the set Q. We assume that the IP-layer always takes the shortest path protocol based on the link-state information as its routing protocol.

Copyright ⓒ 2014 KSII

We have implemented our proposed algorithm by MATLAB and CPLEX [29]. For each simulation scenario, we run the simulation 2000 times and obtain the average value for each performance metric.

5.2 Performance Metrics

During the simulation, we use two performance metrics to evaluate the performance of our proposed algorithm. The first metric is the performance gain for LB-OOMR algorithm:

$$\text{GAIN} = \frac{S_{LB-OOMR}}{S_{non}} \tag{17}$$

For RSM and SND, $\text{GAIN} = S_{RSM}/S_{non}$ and $\text{GAIN} = S_{SND}/S_{non}$, respectively. Larger value of GAIN means smaller congestion ratio and greater network throughput.

To evaluate the reliability of our proposed algorithm, we take Average Link Overlap Ratio as the second performance metric. We define the link overlap of two paths as the number of shared physical links between these two paths. Thus, average link overlap ratio is the average pairwise overlap between a set of k +1 paths over the number of links in the default physical path, in which the set of k +1 paths consists of one default physical path from a source node to a destination node and k one-hop overlay paths between the same source-destination pair. To some extent, smaller average link overlap ratio implies larger path diversity, which is essential to assure the reliability of one-hop overlay routing.

5.3 Simulation Analysis

5.3.1 The Effect of Overlay Network Size

In this section, we analyze the performance of our proposed algorithm under the network topology CN070 and GT180. We set k =4, M =15 in CN070 and M =20 in GT180.

Fig. 6 and **Fig. 7** show the effect of overlay network size on GAIN under CN070 and GT180, respectively. From **Fig. 6** and **Fig. 7**, we can obtain that the value of GAIN obtained by LB-OOMR is significantly greater than that obtained by RSM and SND. Specifically, as shown in **Fig. 6**, LB-OOMR outperforms RSM and SND with around 17% and 13% under CN070, respectively. And **Fig. 7** shows that LB-OOMR achieves almost 60% and 55% greater GAIN than RSM and SND, respectively. These indicate that the congestion ratio obtained by LB-OOMR is smaller than that by RSM and SND.

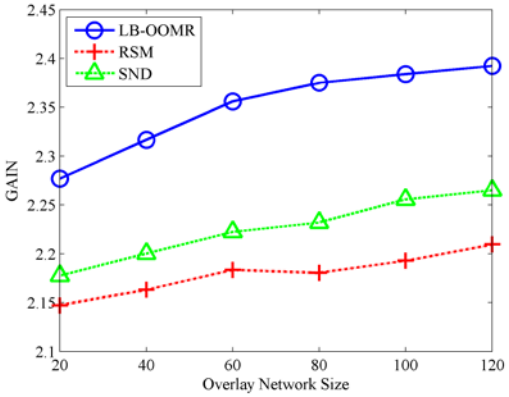


Fig. 6. Overlay network size vs. GAIN under CN070

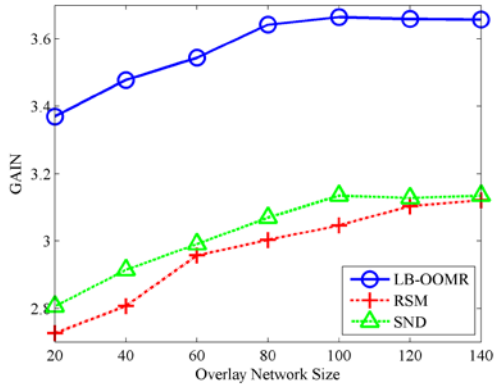


Fig. 7. Overlay network size vs. GAIN under GT180

In addition, the change trend of GAIN obtained by LB-OOMR is similar to that obtained by RSM and SND. GAIN increases as the overlay network size increases. This is because more good nodes are selected as the relay nodes with the increase of overlay network size. Especially for LB-OOMR, with the increase of overlay network size, GAIN under both CN070 and GT180 increases rapidly at the beginning, and then shows a slow increased tendency. This is because the overlay network size can affect the selection of relay nodes. And when the number of overlay nodes is small, a few nodes with higher betweenness centrality are selected frequently as the relay nodes, which results in the link overlap among k one-hop overlay paths, thus increasing the congestion ratio of the network.

Fig. 8 and **Fig. 9** show the effect of overlay network size on Average Link Overlap Ratio under two different topologies: CN070 and GT180. For all three different algorithms LB-OOMR, RSM and SND, we obtain the same result that the average link overlap ratio decreases slightly at first, and then changes smoothly. This is because a larger number of overlay nodes allows more choices of relay nodes, and thus produces better disjoint paths than a configuration with fewer overlay nodes. In addition, from **Fig. 8** and **Fig. 9**, an important observation to make is that the average link overlap ratio obtained by LB-OOMR is far superior to that by RSM and SND regardless of the overlay network size. Specifically, LB-OOMR outperforms RSM and SND significantly with about 40% and 32% improvement under CN070, and about 23% and 30% improvement under GT180. This indicates that LB-OOMR can improve the reliability of one-hop overlay routing.

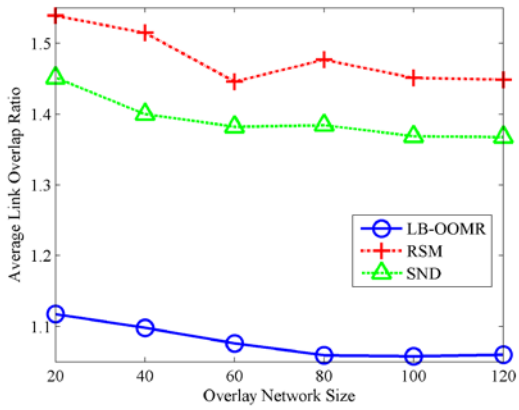


Fig. 8. Overlay network size vs. Link overlap under CN070

Copyright ⓒ 2014 KSII

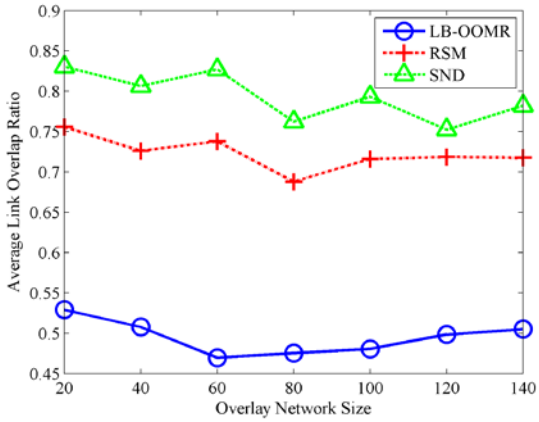


Fig. 9. Overlay network size vs. Link overlap under GT180

5.3.2 The Effect of Candidate Relay Nodes Size

In this section, we study the effect of the number of candidate relay nodes on the performance of our proposed algorithm in terms of congestion ratio and average link overlap ratio. We select randomly 50 nodes as overlay nodes from CN070 and GT180 respectively and vary the number of candidate relay nodes from 5 to 50. The simulation results are shown in **Fig. 10** and **Fig. 11**. From these two figures, we observe that with the increase of the number of candidate relay nodes, the congestion ratio and the average link overlap ratio decrease rapidly at first, and then decrease rather gradually when the number of candidate relay nodes changes from 15 to 50. We also see that the number of candidate relay nodes "15" in CN070 is an inflection point for both congestion ratio and average link overlap ratio, which corresponds to about 10% of total number of nodes. Similarly, "20" is the inflection point in GT180. From these results, we conclude that only a few candidate relay nodes can improve the load balancing and the reliability for one-hop overlay routing. Meanwhile, the fewer number of the candidate relay nodes, the lower complexity of the proposed LP algorithm. In a word, our proposed algorithm is feasible and effective.

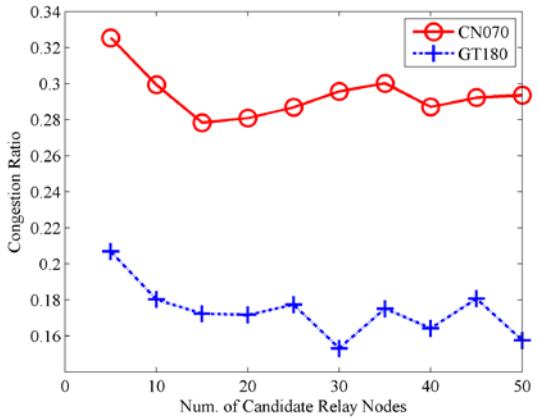


Fig. 10. Num. of candidate relay nodes vs. congestion ratio

In **Fig. 10** and **Fig. 11**, there is a gap between two network topologies (CN070 and GT180) in terms of both congestion ratio and average link overlap ratio. The reason can be explained as follows. In [30], the authors discovered that in the Internet the node degree distribution

follows a power law. CN070 is a real Internet topology that records the interconnection situation of most routers in China in 2006. The degrees of nodes in CN070 are not uniform and there exist a few "core nodes" with large degrees. Note that in LB-OOMR the traffic is rerouted from the source to the relay nodes and from the relay nodes to the destination along the shortest path. These core nodes might be on the shortest paths with high probability, even chosen as the relay nodes, which may increase the congestion ratio and the average link overlap ratio. On the other hand, GT180 is random network topology generated by Waxman model, in which nodes are distributed uniformly in the plane and edges are added according to probabilities that depend on the distances between the nodes. Therefore, each node in GT180 is selected as a relay node with equal probability, which leads to less overlap links among the different paths and lower congestion ratio.

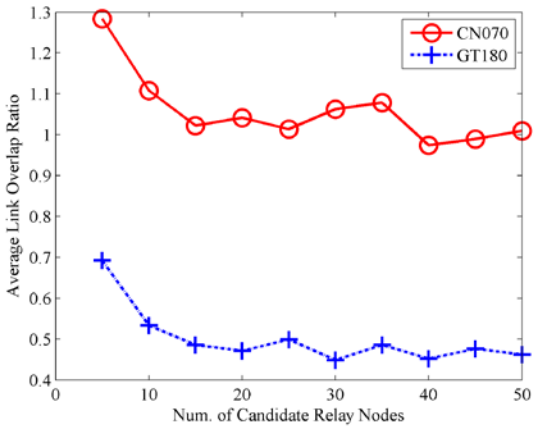


Fig. 11. Num. of candidate relay nodes vs. Average link overlap ratio

5. Conclusion

In this paper, a one-hop overlay multipath routing scheme (LB-OOMR) is addressed by taking into account the load balancing and the path diversity. In our proposed scheme, when a path fails, the source splits the traffic and reroutes them to the destination along multiple one-hop overlay disjoint paths that are established by using a collection of relay nodes. LB-OOMR provides load balancing at the application layer instead of IP layer, which decreases the network overhead and improves the network utilization. To determine a set of optimum split ratios for load balancing, an LP formulation is derived, which is solved with a heuristic algorithm. The simulation results show that our proposed algorithm is fundamentally more efficient in reducing the congestion ratio and improving the reliability of the network.

References

[1] A. Markopoulou, G. Iannaccone, S. Bhattacharyya, C.-N. Chuah, and C. Diot, "Characterization of failures in an IP backbone," in *Proc. of the 23th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, vol. 4, pp. 2307-2317, March 7-11, 2004. Article (CrossRef Link).

[2] M. Venkataraman and M. Chatterjee, "Quantifying video-QoE degradations of Internet links," *IEEE/ACM Transaction on Networking (TON)*, vol. 20, no. 2, pp. 396-407, 2012. Article (CrossRef Link).

Copyright ⓒ 2014 KSII

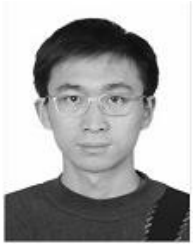
[3] T. G. Griffin and B. J. Premore, "An experimental analysis of BGP convergence time," in *Proc. of the Ninth International Conference on Network Protocols (ICNP)*, pp. 53-61, November 11-14, 2001. Article (CrossRef Link).
[4] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian, "Delayed Internet routing convergence," *IEEE/ACM Transactions on Networking (TON)*, vol. 9, no. 3, pp. 293-306, 2001. Article (CrossRef Link).
[5] C. Boutremans, G. Iannaccone, and C. Diot, "Impact of link failures on VoIP performance," in *Proc. of Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, May 12-14, 2002. Article (CrossRef Link).
[6] D. Andersen, H. Balakrishnan, F. Kaashoek, and R. Morris, "Resilient overlay network," in *Proc. of ACM Symposium on Operating Systems Principles (SOSP)*, pp. 131-145, 2001. Article (CrossRef Link).
[7] X. Zhou, D. Guo, T. Chen and X. Luo, "Robust backup path selection in overlay routing with bloom filters," *Transactions on Internet and Information Systems (KSII)*, pp. 1890-1910, 2013.
[8] Z. Li and P. Mohapatra, "QRON: QoS-aware routing in overlay networks," *IEEE Journal on Selected Areas in Communications (JSAC)* , vol. 22, no. 1, pp. 29-40, 2004. Article (CrossRef Link).
[9] K. P. Gummadi, H. Madhyastha, S. D. Gribble, H. M. Levy, and D. J. Wetherall, "Improving the reliability of internet paths with one-hop source routing," in *Proc. of Symposium on Operating Systems Design and Implementation (OSDI)*, 2004.
[10] R. Cohen, and D. Raz, "Cost effective resource allocation of overlay routing relay nodes," *IEEE/ACM Transactions on Networking (TON)*, 2013. Article (CrossRef Link).
[11] S. Roy, H. Pucha, Z. Zhang, Y. C. Hu, and L. Qiu, "On the placement of infrastructure overlay nodes," *IEEE/ACM Transactions on Networking (TON)* , vol. 17, no. 4, pp. 1298-1311, 2009. Article (CrossRef Link).
[12] R. Kawahara, S. Kamer, N. Kamiyama, H. Hasegawa, H. Yoshino, Eng Keong Lua and A. Nakao, "A method of constructing QoS overlay network and its evaluation," in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM)*, 2009. Article (CrossRef Link).
[13] S. Savage, A. Collins, E. Hoffman, J. Snell, and T. Anderson, "the end-to-end effects of internet path selection," in *Proc. of ACM SIGCOM*, pp. 289-299, 1999. Article (CrossRef Link).
[14] M. Cha, S. Moon, C. D. Park, and A. Shaikh, "Placing relay nodes for intra-domain path diversity," in *Proc. of the 25th IEEE International Conference on Computer Communications (INFOCOM)*, April 2006. Article (CrossRef Link).
[15] J. Liao, J. Wang, B. Wu, and W. Wu, "Toward a multiplane framework of NGSON: a required guideline to achieve pervasive services and efficient resource utilization," *IEEE Communications Magazine*, vol. 50, no. 1, pp. 90-97, 2012. Article (CrossRef Link).
[16] G. Lee and J. Choi, "A survey of multipath routing for traffic engineering," 2002. [Online]. Available: http://vega.icu.ac.kr/~gmlee/research/.
[17] R. K. Singh, N. S. Chaudhari, and K. Saxena, "Load balancing in IP/MPLS networks: a survey," *Communications and Networks*, vol. 4, pp. 151-156, 2012. Article (CrossRef Link).
[18] Y. Yoshida and M. Kawarasaki, "Relay-node based proactive load balancing method in MPLS network with service differentiation," in *Proc. of IEEE International Conference on Communications (ICC)*, pp.7050-7054, 2012. Article (CrossRef Link).
[19] E. Oki and A. Iwaki, "Load-balanced IP routing scheme based on shortest paths in hose model," *IEEE Transactions on Communications (TOC)* , vol. 58, no. 7, pp. 2088-2096, 2010. Article (CrossRef Link).
[20] U. Brand, "On variants of shortest-path betweenness centrality and their genetic computation," *Social Networks*, vol. 30, no. 2, pp.136-145, 2008. Article (CrossRef Link).
[21] G. Zhang, "An algorithm for Internet AS graph betweenness centrality based on Backtrack," *Journal of Computer Research and Development*, vol. 40, no. 10, pp. 1790-1796, 2006. Article (CrossRef Link).
[22] V. Padmanabhan, L. Qiu, and H. Wang, "Server-based inference of Internet link lossiness," in *Proc. of the 22th Annual Joint Conference of the IEEE Computer and Communications Societies*

(INFOCOM), vol. 1, no. 1, pp. 145-155, 2003. Article (CrossRef Link).
[23] A. Nakao, L. Peterson, and A. Bavier, "A routing underlay for Overlay Networks," in *Proc. of ACM SIGCOMM*, 2003. Article (CrossRef Link).
[24] V. Aggarwal, A. Feldmannn, and C. Scheideler, "Can ISPs and P2P users cooperate for improved performance?" *ACM SIGCOMM Computer Commun. Rev. (CCR)*, vol. 37, no. 3, pp. 29-40, 2007. Article (CrossRef Link).
[25] H. Xie, Y. R. Yang, A. Krishnamurthy, Y. Liu and A. Silberschatz, "P4P: provider portal for applications," in *Proc. of ACM SIGCOM*, 2008. Article (CrossRef Link).
[26] K. Tutschku, T. Zinner, A. Nakao, and P. Tran-Gia, "Network virtualization: Implementation steps towards the future internet," in *Proc. of the Workshop on Overlay and Network Virtualization at KiVS*, March 2009.
[27] GT-ITM: Modeling Topology of Large Internetworks [Online]. Available: http://www.cc.gatech.edu/projects/GT/.
[28] B. M. Waxman, "Routing of multipoint connections," *IEEE Journal on Selected Areas in Communication (JSAC)* , vol. 6, no. 9, pp. 1617-1622, 1988. Article (CrossRef Link).
[29] ILOG, Inc, "ILOG CPLEX: High-performance software for mathematical programming and optimization," 2006, Available: http://www.ilog.com/products/cplex/.
[30] M. Faloutsos, P. Faloutsos, C. Faloutsos, "On power-law relationships of the Internet topology," *ACM SIGCOMM Computer Commun. Rev. (CCR)*, vol. 29, no. 4, pp. 251-262, 1999. Article (CrossRef Link).

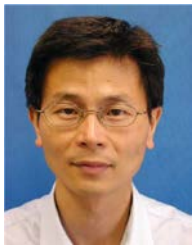
Copyright ⓒ 2014 KSII

Jianxin Liao was born in 1965, obtained his PhD degree at University of Electronics Science and Technology of China in 1996. He is presently a professor of Beijing University of Posts and Telecommunications. He has published hundreds of papers in different journals and conferences. His research interests are mobile intelligent network, broadband intelligent network and 3G core networks. He is the Specially-invited Professor of the "Yangtse River Scholar Award Program" by the China Ministry of Education in 2009.

Shengwen Tian was born in 1974, obtained his BS degree in computer science and technology from Qingdao University in 2005. He is currently working toward the PhD degree in computer science and technology at Beijing University of Posts and Telecommunications, China. Now he is a lecturer in Ludong University, China. His research interests include overlay networks, Next Generation Network, complex networks, and data mining.



Jingyu Wang was born in 1978, obtained his PhD degree from Beijing University of Posts and Telecommunications in 2008. Now he is an associate professor in Beijing University of Posts and Telecommunications, China. His research interests span broad aspects of performance evaluation for Internet and overlay network, traffic engineering, image/video coding, multimedia communication over wireless network.



Tonghong Li was born in 1968, obtained his PhD degree from Beijing University of Posts and Telecommunications in 1999. He is currently an assistant professor with the department of computer science, Technical University of Madrid, Spain. His main research interests include resource management, distributed system, middleware, wireless networks, and sensor networks.

Qi Qi was born in 1982, obtained his PhD degree from Beijing University of Posts and Telecommunications in 2010. Now she is an assistant professor in Beijing University of Posts and Telecommunications. Her research interests include SIP protocol, communications software, Next Generation Network, Ubiquitous services, and multimedia communication.