A Measurement Study on the Benefits of Open Routers for Overlay Routing

Li Tang[*,a,c] Yin Huai[a] Jin Zhou[c] Hao Yin[b,c] Zhen Chen[c,d] Jun Li[c,d]
[a] Department of Automation, Tsinghua University, Beijing, 100084, China
[b] Department of Computer Science, Tsinghua University, Beijing, 100084, China
[c] Research Institute of Information Technology, Tsinghua University, Beijing, 100084, China
[d] Tsinghua National Lab for Information Science and Technology, Beijing, 100084, China
Email: {tangli03, huaiy05}@mails.tsinghua.edu.cn {zhoujin, h-yin, zhenchen, junl}@tsinghua.edu.cn

Abstract— **As the rapid proliferation of multimedia communications has been significantly challenging the Internet on providing a high-quality transporting service to achieve desirable audio and video user experience, overlay routing is proposed as a practical technique to complement some inherent deficiencies of the current IP routing service. While many research and realistic systems have proven the effectiveness of overlay routing on improving end-to-end (E2E) performance, this paper performs a measurement study on the potential benefits of *open router*s, a novel promising technical concept emerging recently, to boost overlay routing for multimedia communications. Based on analysis of a large amount of measurements collected between more than 200 PlanetLab sites over the world, our results show that when being deployed appropriately, even a small number of open routers will be able to remarkably enhance overlay routing's ability of improving E2E performance for multimedia communications. Accordingly, several practical heuristics for selecting suitable places to deploy open routers are proposed and comparatively evaluated based on real Internet topological and routing data.**

Index Terms— **overlay routing, open routers, multimedia communications, heuristic deployment**

I. Introduction

Recently, worldwide multimedia communications that aim to provide users with racy and lifelike experience by making use of audio and video techniques are becoming another popular type of applications on the Internet. According to Cisco forecast, video is now approximately one-quarter of all consumer Internet traffic and will increase to one-half by 2012 [1]. On the other hand, the fundamental architecture and design principles of the Internet mostly remain the same as they were originally created just for the data transmission purpose. As a result, the high interactivity and real-time requirement of novel multimedia communications are triggering significant challenges to the current Internet's IP routing service. However, due to the vast investment of changing IP infrastructures and the lack of coordination between ISPs, so far there is still no QoS solution universally deployed across the Internet. Instead, overlay routing techniques have been widely used in practice to overcome inherent limitations of the Internet's best-effort service [2], [3], [4], [5], [6].

This work is supported by NSFC grants No.90718040 and NEC Laboratories China.

Manuscript received March 1, 2009; revised June 12, 2009; accepted June 22, 2009.

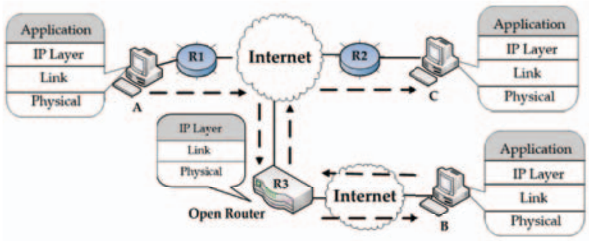


Figure 1. An example use of open routers

The key idea of overlay routing is to better fit application-specific requirements by using proper overlay paths besides the single IP path provisioned by Internet's IP routing service itself. In existing overlay routing systems, an overlay path consists of a series of relays that forward packets on application layer. Undoubtedly, to what extent overlay routing is able to detour degraded IP paths and thus enhance E2E performance depends on how many and how appropriate relays are available for use. Therefore, most existing systems are subjected to their lack of ability to access and control routers, and can only utilize end-hosts located in edge networks to play the role of relays. Limiting overlay routing to application layer and relays to end-hosts can lead to unnecessary inefficiency. As an example shown in Figure 1, if an end-host *B* located in edge networks is selected to be a relay, the data transmitted on the corresponding overlay path have to fruitlessly travel through *B*'s edge networks twice (once ingress and once egress).

In recent years, in order to overcome the impasse of Internet architecture, *virtualization* has been proposed and advanced as a promising means to reinvigorate architectural innovation that can be practically applied on the current Internet [7], [8], [9], [10]. Although the future Internet's architecture is still not lucid, we consider such a technical renewal as a valuable opportunity to remedy current overlay routing's deficiencies. At the very least, had the router *R3* been able to act as the relay instead of *E* in the given example, both the ingress and egress transmission of *E*'s edge networks could be avoided,

doi:10.4304/jcm.4.9.714-723

which would reduce not only the packet's E2E delay but also its loss probability. With such observations, this paper aims to investigate if some routers in the Internet really become *virtualized* in future, whether and how overlay routing would benefit from this renovation. We refer to a virtualized router that can play the role of a relay as an *open router*, a term named by one of the main urges of Internet virtualization, Jonathan Turner [9], [10], [11].

The focus of this paper is not to debate open routers will surely come off in future, but instead, taking it as a promising possibility, we aim to reveal to what extent overlay routing can benefit from the emergence of open routers. Our main contributions include:

• To the best of our knowledge, this work is among the first research efforts investigating many potential advantages that open router can bring to overlay routing for improving multimedia communications.
• In order to evaluate the advantages quantitatively, a research methodology is developed to conduct an experiment to collect a large amount of measurements between more than 200 different PlanetLab sites and compare the effects of overlay routing on improving E2E performance with and without leveraging open routers.
• Considering that open routers are likely to be gradually deployed, a variety of topological heuristics that may be practically used to select suitable deploying places of open routers are proposed and evaluated.
• Last but not least, most of the techniques we used to collect, process and formalize the measurement data can be easily generalized or extended, to investigate similar research topics regarding to the Internet's topological and routing natures, such as the server placement problem in CDN networks.

II. Background and Motivation

The Internet architecture was originally developed to achieve a set of different goals with ordered priorities. However, over years since the Internet started, many situations that used to be the basis of some architectural principles have vastly changed. In this section, we present the background and motivation of our study on three aspects.

A. Stringent Requirements of Multimedia Communications

In the past decade, many people started to use the Internet as a platform for their multimedia communications. Not only customers hope to enjoy communicating lively and economically, but also service providers have their own incentives to make profits from new applications inspired by multimedia communications.

In the Internet's original design, its elementary function is to provide a basic packet delivery service between two end-hosts if they are connected on physical layer. Such a 'best-effort' service model accomplishes simplicity and compatibility, and helps the Internet become a great success. However, the best-effort service is no longer able to satisfy the requirements of multimedia communications, where achieving a proper performance level for each packet becomes a fundamental part of the delivery service. In order to attain an agreeable user experience in communications, it is necessary to make the expected audio or video packet delivered not only successfully but also timely. Therefore, deliberate retransmission mechanisms, as adopted in TCP or RTP/RTCP, become useless, because the retransmitted multimedia data are likely to be too stale to play out. Although RTP protocol has been designed to suitably transmit real-time data, it does not provide any mechanisms to ensure QoS guarantees, but depends on lower-layer service to do so. In this sense, RTP transmits packets for multimedia communications through the same route across the Internet as that used by data transmission applications.

Moreover, multimedia packets often carry information of different significances and priorities and thus have different QoS requirements. Given this, QoS guarantee can be implemented either based on flows by over-provisioning each flow to meet the most stringent requirement of all its packets, or based on packets but taking the risk of violating canonical architectural principles, vastly complicating the router's implementation, and interfering other applications that do not need QoS at all. Both measures have negative effects on making profits and are not favored by ISPs.

B. Rise of Overlay Routing and Open Routers

Given too many difficulties in implementing QoS on IP layer, overlay routing has been successfully applied in improving the quality of multimedia communications [2], [4], [5] due to several reasons. First, the current Internet's IP routing comprises remarkable inflation of E2E performance [12], and in many cases there are alternate overlay paths superior to the direct IP path [13], [14], [3]. Second, overlay routing supports application-specific routing, because it can construct multiple overlay paths with diverse E2E performance, and transmit each packet independently through the most suitable overlay paths according to the packet's application-specific requirement. Last, overlay routing does not necessitate remarkably change of the Internet's legacy routing infrastructure, which therefore makes it much easier to implement and to deploy in practice.

There has been a long debate between architectural purists and pluralists on preferable architecture design for the next generation Internet. *Virtualization* has been proposed as a promising measure to break the impasse and advance innovations [7], [8], [9]. However, our position in this paper is neither to argue that open routers will definitely become realized in future, nor to propose a comprehensive solution of the open router's design, implementation or deployment. Instead, we take open routers as a promising technical concept in the interim while the current Internet is evolving towards its successor, and aim to reveal an objective evaluation on whether and to what extent open routers would be able to benefit overlay routing. Our study intends to help researchers in charge of designing transmission mechanisms for multimedia communications gain a clear understanding on the impact of the possibly oncoming open routers, which can help them base their design and use of overlay routing techniques on a long-term insight. The detailed discussion on the open router's design, implementation or deployment in the future Internet's architecture is beyond the scope of this paper; readers interested in such topics may refer to [7], [8], [9].

C. Integrating Overlay Routing and Open Routers

We particularlly concentrate on overlay routing techniques that use one-hop source routing [15], where each overlay path is composed of a single relay, due to the following two main reasons. For one thing, it has been widely indicated most performance gains of overlay routing can be obtained by utilizing just a single relay [14], [15], [3]. For another, the design leaves intelligence to the source end-host, and keeps the simplicity of networks, which is widely acknowledged as a key reason for Internet's success. Such a design minimizes the complexity of the relay's implementation, because the routing information of an overlay path can be encoded into the packet's header by the source end-host and all that a relay needs for acting a relay is to simply forward the received packet according to the destination's identifier encoded in the packet's header.

Based on one-hop source routing, overlay routing techniques can be taken as an abstract framework that can intelligently manipulate a set of overlay paths. The source end-host maintains every overlay path's performance, and delivers a packet through one or multiple suitable overlay paths that best fit the packet's application-specific requirement. Each open router that is integrated into the overlay routing system runs a virtualized router that functions as a relay. Conceptually, an open router can be considered as an infrastructural relay node that is inherently implemented in a router and placed in the core of networks.

D. Potential Advantages of Integrated Design

Based on the above design, we perceive a variety of potential advantages that open routers can bring to overlay routing. First, open routers can help to construct high-quality overlay paths by eliminating transmission in edge networks. Without open routers, overlay routing can only use end-hosts to play the role of relays. As the bandwidth of the end-host's access networks is generally smaller than that of the Internet's backbone, it often becomes the whole overlay path's performance bottleneck. In contrast, using the relay played by an open router can save the round trip between the open router and previous end-host-relay; therefore, it reduces the overlay path's E2E delay and the overlap of its underlying IP paths.

Second, open routers ease the measurement of performance on IP layer. Because open routers also run other virtualized routers undertaking the common IP router's functionality, they can monitor the traffic passing through in a passive manner, to explicitly provide accurate performance information of its directly connected links. Moreover, open routers are able to directly obtain some under-layer information, such as the link's capacity, which is extremely difficult to measure accurately with E2E techniques on application layer.

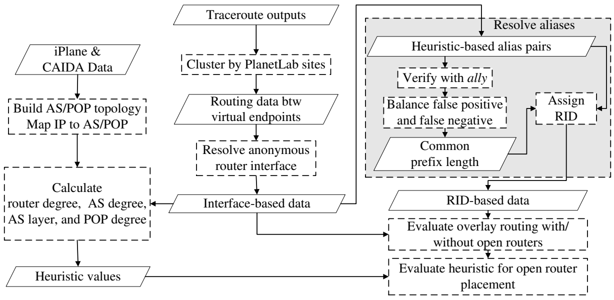


Figure 2. Flow process of our research methodology

Finally, open routers themselves are infrastructural devices much more reliable and stable than another type of relays played by common user's end-hosts. The high churn rate of end-host-relays significantly complicates the algorithm for selecting suitable overlay paths. Even worse, if an overlay path is being used to transmit critical data in multimedia communications, its interrupt caused by the corresponding relay's off-line or collapse will severely degrade the user's satisfaction. Because it takes time for the source end-host to realize the interrupt before taking action to switch to another path, all traffic supposed to be transmitted by the quondam overlay path will be dropped during the period.

III. Measurement and Analysis Methodology

In order to quantitatively study the benefits of open routers on overlay routing, the key challenge of this paper is to design a methodology that can obtain the Internet's real-life topology and routing information on the router level with the currently available measurement approaches, given that there have not been realistic open routers deployed in the Internet core so far. In this section, we first define two metrics to evaluate overlay routing's ability of improving the primary E2E performance factors that multimedia communications are sensitive to, and then explain the whole flow process of our research methodology as sketched out in Figure 2.

A. Evaluation Metrics

Given a one-hop overlay path constituted by a relay r between a pair of source s and destination d, we define the following two metrics for evaluating the overlay path's potential to assist the transmission of multimedia data:

◇ **E2E-delay**: the sum of the delay between s and r and the delay between r and d, i.e., $D_{srd} = D_{sr} + D_{rd}$, where all variants are measured by RTT in practice.

◇ **Overlap**: the overlay path's *overlap* to its corresponding direct IP path is the sum of *overlap values* between every two out of the three direct IP paths *sr*, *rd*, and *sd*. The *overlap value* between two direct IP paths $P_1 = (a_1, a_2, \ldots, a_m)$ and $P_2 = (b_1, b_2, \ldots, b_n)$ are defined to be $Overlap(P_1, P_2) = Nodes(P_1) \otimes Nodes(P_2) - 1 + Links(P_1) \otimes Links(P_2) = \sum_i \sum_j a_i \oplus b_j + \sum_k \sum_l \overrightarrow{a_k a_{k+1}} \oplus \overrightarrow{b_l b_{l+1}} - 1$, where *Nodes(P)* and *Links(P)* respectively stand for the functions mapping an IP path *P* to its nodes and links, and $\oplus$ denotes an operator defined as $x \oplus y = 1$ if $x = y$, or 0 otherwise.

The first metric characterizes the delay-critical feature of multimedia communications, while the second metric represents the performance correlation between the overlay path and its corresponding IP path. Therefore, for either metric, the smaller it is, the larger potential the overlay path possesses to assist the transmission of multimedia data. Similar metrics have been used for evaluating and selecting overlay paths in previous research [16].

B. Data Collection

In order to characterize the Internet's heterogeneity and topological diversity, we deployed a set of scripts on every available PlanetLab node to conduct them to `traceroute` each other. We left all parameters of `traceroute` to be default, except increasing the maximum time-to-live (TTL) value to 255 (enough for the Internet's current scale) to avoid the unreachability to a destination caused by TTL limit. From each raw output of a run of `traceroute`, we extracted two types of information: an E2E path consisting of a sequence of hops, each of which is represented by corresponding router interface's IP address; and the E2E delay measured by the minimum RTT obtained by relevant probes from source to each hop along the path.

We collected nearly 300000 `traceroute` outputs in total during the week starting since March 16, 2008. Then, we filtered out `traceroute` outputs failing to reach destinations as well as those containing routing loops or multiple contiguous anonymous interfaces, because all of them could not provide intact information to our study. The remained `traceroute` outputs are between 199790 PlanetLab node pairs covering totally 574 different nodes.

Usually, each PlanetLab site hosts multiple nodes in the same subnet. As our primary focus is on topological and routing issues across the Internet, we group all nodes belonging to the same site into one cluster and mapped them into a *virtual endpoint*. Conceptually, a virtual endpoint corresponds to an end-host in real applications of multimedia communications, and in this paper we particularly focus on routing characteristics between different virtual endpoints. In following discussion we often use 'endpoint' instead of 'virtual endpoint' for briefness. For each pair of endpoints, the `traceroute` output with the most intact information is selected out of all available ones with their source and destination PlanetLab nodes respectively belong to the corresponding two PlanetLab sites, to be the *E2E path* between this pair of endpoints. Retrospectively, this redundant collection of `traceroute` outputs has indeed alleviated the interference of accidental failures of PlanetLab nodes, and significantly improved the coverage of the constructed network topology and the integrality of routing information between endpoints.

After above pretreatments, the resulted data set consists of E2E paths between 46578 pairs of endpoints, and totally covers 259 (218 source and 249 destination) different endpoints, i.e. PlanetLab sites. Given that these PlanetLab sites are located in 177 different autonomous systems (ASes) and geographically distributed over the world, as shown in Appendix A, it is convincing that the data set in our study can plausibly represent a rich cross-section of situations in present Internet-scale overlay networks.

C. Router Interface Resolution

So far, all routing information in E2E paths is based on router interfaces. As a physical router usually has multiple interfaces, each of which is assigned an individual IP address, obtaining the router-level topology and routing information of the generated overlay networks has to appropriately resolve and aggregate interfaces that belong to the same router. To solve this problem, we develop a three-step method to map a router interface's IP address to the corresponding physical router's identity. To elaborate our methods clearly, we define several terms as follows:

◇ **RID**: A *RID* is the unique identity of a node in the inferred router-level topology graph and is supposed to stand for a physical router on the Internet in practice.

◇ **Alias**: An *alias* of a router is an IP address or an asterisk that represents a specific interface of this router in an E2E path, i.e., `traceroute` output. The aliases of a router appearing in all relevant E2E paths compose this router's *alias set*.

◇ **Loop**: An E2E path contains a loop, if and only if the path contains multiple hops with the same identity (IP or RID).

◇ **Position**: Given a path consisting of a sequence of hops between a pair of source *s* and destination *d*, the *position* of a hop *h* on this path is a directed 2-tuple consisting of the distance from *s* to *h* and the distance from *h* to *d*, where the distance is measured by the number of hops.

The first step of our router interface resolution method is to deal with anonymous interfaces represented by '$*$' in `traceroute` outputs. While sophisticated techniques have been proposed [17], the difficulty in selectively using these techniques to fit our interests is how to make a suitable trade-off between false positive (wrongly attributing interfaces of different routers into the same router ID) rate and false negative (missing discovering a router's actual interfaces) rate. Generally, a high false positive rate tends to cause an overestimation of the open router's effect, while a high false negative rate will generate the opposite side effect. Because there are other inevitable factors making open router's effect underestimated, as analyzed in the next section, we give higher priority to reducing the false positive rate in this step; otherwise, it would be difficult to analyze the impact of systematic errors on the open router's actual effect. Specifically, for each anonymous interface, if there exists and only exists one non-anonymous interface having the same predecessor and successor interfaces as this anonymous one in all relevant E2E paths, we consider they are identical to each other.

In the second step, we attempt to discover every router's potential alias set. One publicly available source of alias sets comes from the iPlane project [18], as part of the service to build an information plane for Internet-scale distributed systems. Using the same-returned-source-address and link-ends-with-common-prefix heuristics, iPlane identifies alias candidates among the interfaces observed in its extensive `traceroute` measurements. Afterwards, iPlane confirms true aliases out of alias candidates as what returning similar IP-IDs and identical TTLs when probed at the same time.

Despite a good source, iPlane does not necessarily observe every interface appearing in our interface-based data and its list of alias sets refreshes infrequently. To complement the deficiencies, we exploit two other heuristics based on the current Internet's routing behavior to further identify alias candidates in our topology. The first heuristic is routing symmetry, and the other heuristic is based on the destination-dependent packet forwarding, i.e., if two interfaces are in two E2E paths with the same destination and have the same predecessor, they are likely to be aliases of the same router. Appendix B presents pseudo-codes implementing these two heuristics.

Afterwards, we filter out every alias pair causing loops in one or more E2E paths. In order to alleviate false positives, we further verify the common prefix of each remained alias pair: we judge a pair of alias candidates to be true aliases only if they both have public IP addresses and the length of their common prefix is larger than N. Similarly, deciding a proper value of N also involves the trade-off between false positive and false negative in the alias identification result. To solve this problem, we leverage the *ally* tool, an alias resolver currently implemented in Scriptroute system [19], to verify the validity of candidate alias pairs identified by our heuristics. *Ally* returns explicit results consistently to around 43% candidate alias pairs, which suggest 18 (Appendix C explains our detailed approach) be the best value of N to verify the other 57% candidate alias pairs that *ally* cannot judge their validities. Ultimately, the verified alias pairs contain 7120 unique interfaces with public IP addresses, which are clustered into 1559 different alias sets (routers).

In the last step, we attribute each router interface to its corresponding router's RID. To achieve this, we first group the interfaces with public IP addresses into clusters based on all alias sets provided by iPlane and generated by the techniques described in the last paragraph. Afterwards, every interface in the same cluster is attributed to the same RID as long as no loop is resulted in any E2E path. If attributing two interfaces to an identical RID causes a loop, both interfaces will be excluded from their cluster and each will be attributed to an individual RID different from the cluster's one.

Interfaces having no public IP address are either anonymous or ones with private IP addresses. As a private IP address does not guarantee global uniqueness and may be used by different routers in different routing domains, we cannot simply attribute interfaces with identical private IP address in different E2E paths to the same RID. Neither can we do this to anonymous interfaces. Instead, we attribute two interfaces with the same private IP address or anonymous in different E2E paths to the same RID, only if they have both identical public-IP predecessor and identical public-IP successor, and in addition, they have the same position in corresponding paths from their predecessors to successors. Every remained individual interface is attributed to a distinct RID, as we do to each alias cluster.

The final topology and routing data used for evaluation in the following section totally contain 259 endpoints, 46578 E2E paths (including 771297 hops), and 6430 routers (aggregated from 10810 interfaces) scattered in 705 different ASes.

IV. Effect of Open Routers

In this section, we aim to investigate whether and to what extent the deployment of open routers can enhance overlay routing to benefit multimedia communications. Specifically, based on the router-level topology and routing data generated above, we compare the statistical E2E-delays and overlaps of the overlay paths respectively with and without routers being allowed to act as open routers. We begin with assuming that open routers have been universally deployed on the Internet, and then we inspect a more practical circumstance in which open routers are considered expensive and will be increasingly deployed on partial routers.

A. Universally Deployed Open Routers

Theoretically, any intermediate network device with the open router deployed can act as a relay for every pair of end hosts, since the Internet offers a routing service to almost any publicly routable IP address. However, in case of this paper our study is subject to the topological and routing information that we can obtain from the current Internet with existing measurement approaches. As a result, a router, represented by a non-leaf node r in the topology graph, is included in our statistics as a relay for a pair of endpoints s and d, if and only if the routing information from s to r and that from r to d can both be extracted from some E2E paths in our routing data. Note this is an inherent limitation in measuring the Internet's topology and cannot be completely eliminated, because it is almost impossible to discover every link and router on the Internet. Such an impact may lead to a little underestimation of open router's effect.

Figure 3 illustrates the cumulative distribution function (CDF) plots of E2E-delays between every pair of endpoints in a variety of scenarios. In each scenario, all available overlay paths between each pair of endpoints

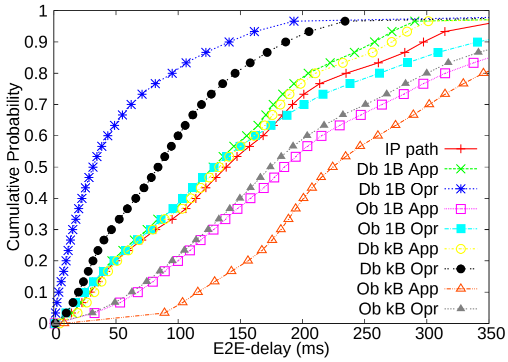


Figure 3. CDF plots of E2E-delays between every pair of endpoints

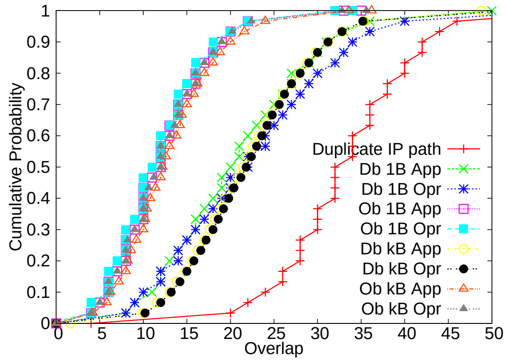


Figure 4. CDF plots of overlaps between every pair of endpoints

are sorted according to a specific metric, and then we do statistics on the E2E-delay of the best overlay path. To alleviate possible biases caused by the single best overlay path, we also calculate the average E2E-delay of the best k (in case of presented plots, k is assigned to be 5) overlay paths for each pair of endpoints. However, due to space limit, we often only present and refer to the best overlay path's characteristics in following discussion, unless there are meaningful observations on the difference.

For the sake of conciseness, we use abbreviations to name the legends: 'Db' and 'Ob' indicate which metric, E2E-delay or overlap, is used for sorting the relevant overlay paths; '1B' and 'kB' indicate whether the plot corresponds to statistics of the best path or to the average statistics of the best k paths; and 'App' and 'Opr' indicate whether end-hosts or open routers play the role of relays. For example, the legend of 'Db 1B Opr' is interpreted as the statistics of the best overlay paths sorted based on the E2E-delay metric without using open routers as relays.

As can be seen, when E2E-delay itself is used as the metric to sort all available overlay paths (plots starting with 'Db'), the E2E-delays of selected overlay paths are always statistically superior to those of corresponding direct IP paths. This fact confirms that overlay routing can indeed reduce E2E-delay for multimedia communications. Moreover, comparing counterpart plots respectively ending with 'App' and 'Opr', we can find that overlay paths constituted of end-host-relays have only a little advantage over the direct IP paths, while the use of open routers as relays makes the E2E-delay of constructed overlay paths remarkably smaller. The results not only demonstrate our conjecture that open routers can effectively enhance overlay routing's ability of improving E2E performance, but also indicate that their effect could be significant in terms of reducing E2E-delay, if they were universally deployed.

Similarly, Figure 4 shows the CDF plots of the other metic, overlaps, between every endpoint-pair's relevant overlay path and corresponding direct IP path. As a reference, we also give the CDF plots of overlaps calculated from two duplicate direct IP paths between every pair of endpoints, which in practice can be considered as the situation of transmitting each packet twice through the direct IP path. Compared to reducing E2E-delay, open router's effect on reducing overlap is much smaller. By comparing plots starting with 'Ob', we observe that the selected overlay paths constituted of open routers precedes those constituted of end-host-relays only a little. This is because replacing end-hosts with suitable open routers to play the role of relays, as illustrated in Figure 1, is only able to effectively reduce the overlap inside the overlay path itself, i.e., the part between sr and rd, but has little impact on the other two parts, i.e., the overlaps respectively between sr and sd and between rd and sd; however, between more than 90% pairs of endpoints, the first part merely accounts for less than 10% of the whole overlap values of the overlay paths that have the smallest overlap with corresponding IP paths.

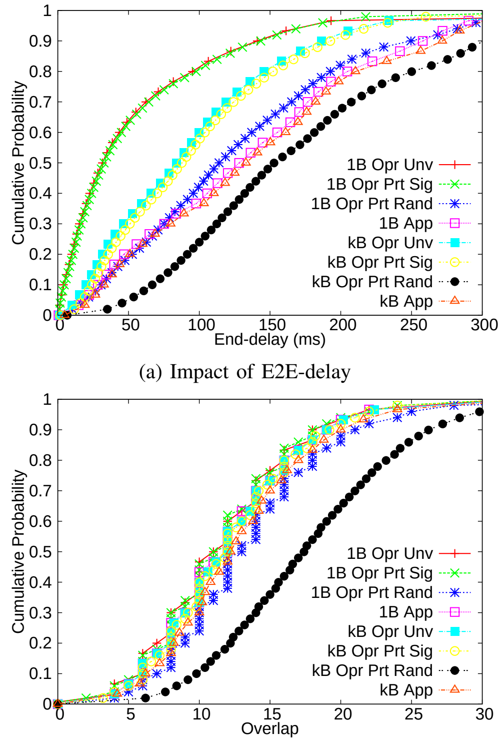
For the sake of comparison, when illustrating statistics of both evaluation metrics, we also present results in which overlay paths are sorted and selected according to the other metric (plots starting with 'Ob' in Figure 3 and with 'Db' in Figure 4). Clearly as it shows, neither metric is perfect: when overlay paths are selected according to one metric, the other metric of the selected overlay paths is inevitably degraded. In real multimedia communications, overlay routing needs to elaborately manipulate multiple overlay paths with different performance characteristics based on application-specific requirements. Despite of this, the ability of constructing overlay paths that can agreeably satisfy different metrics separately, as studied in our methodology, is actually the prerequistite for overlay routing techniques to meet sophisticated requirements and finally improve the quality of multimedia communications.

B. Partially Deployed Open Routers

The last subsection shows that open routers are promising when universally deployed. However, due to cost and security concerns, ISPs may only agree to partially deploy open routers in their networks for sale. It is important to understand open router's effect in such a practical circumstance. However, as elaborated in Appendix D, finding the optimal deployment of a given number of open routers is actually an *NP-hard* problem. Given this, we propose a greedy algorithm that selects a given number of routers with the most significance as preferable deploying places of open routers. The complexity of this algorithm is $O(m^2 \log n)$, where m is the number of all endpoint-pairs, n is the number of routers, and a router's *significance* is defined as follows in our topology:

◇ **Significance**: A router's significance is the proportion of the number of endpoint-pairs for which the router, if with open router deployed, can constitute the best overlay path, in terms of a specific evaluation metric, to the total number of all concerned endpoint-pairs.

Due to the limit of space, we only present a typical scenario where open routers are restricted to deploy on 259 (equivalent to the number of end-host-relays) routers, which are around 4% of all routers observed in our topology data. Figure 5 illustrates the statistics of constituted overlay paths respectively with universally and partially deployed open routers. We note two interesting observations regarding to our proposed algorithm and issues on partial deployment of open routers. First, the statistical E2E-delays and overlaps achieved by using the partially deployed open routers are both very close to those achieved by deploying open routers on every router, indicating that the heuristic of relay significance, as we propose, is effective for selecting suitable deploying places of open routers. Second, given that the optimal placement of open routers should achieve even better performance than that achieved by our algorithm, the result demonstrates that deploying open routers on just a small subset of all routers, if properly selected, can still remarkably enhance overlay routing's ability, but with much lower expenses.



(a) Impact of E2E-delay

(b) Impact on Overlap

Figure 5. Impact of deploying open routers on 4% most significant routers. In the legend, 'Unv' and 'Prt' stand for scenarios where open routers are *universally* and *partially* deployed respectively.

V. Heuristics for Open Router's Placement

While effective as a criteria to guide the deployment of open routers, the router's significance is difficult to use in practice, because it requires the knowledge of routing information between every pair of endpoints that have potential to use the overlay routing service. In this section, we propose several practical heuristics based on the Internet's topological characteristics to select suitable deploying places of open routers:

◇ **Router Degree** (RTD): Prefer routers having larger number of neighboring routers.

◇ **AS Degree** (ASD): Prefer routers located in the ASes having larger number of neighboring ASes.

◇ **AS Layer** (ASL): Prefer routers located in ASes on upper layers in ISP's commercial hierarchy, where layer 0 corresponds to Tier-1 ASes and layer n corresponds to ASes that are n-grade customers of their nearest Tier-1 ASes.

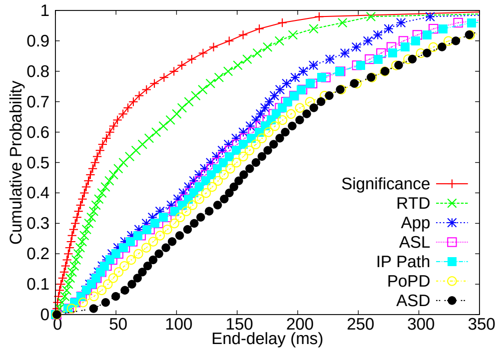
◇ **POP Degree** (POPD): Prefer routers located in POPs having larger number of neighboring POPs.

Given the RID-based topology generated above, it is straightforward to obtain each router's degree. To obtain the other heuristic values of a router, we leverage the publicly available measurement data regarding to the Internet's topological features, which are shared by other research projects, namely iPlane [18] and CAIDA AS Relationships [20].

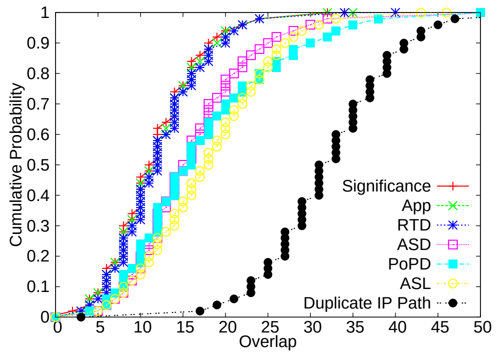
Intuitively, the router having high degree, located in a high-degree or upper-layer AS, or located in a high-degree POP, is likely to have good connectivity and capacity, and can often construct high-performance overlay paths for a variety of endpoint-pairs. Therefore, this type of routers are expected to be preferable deploying places of open routers.

To inspect the effectiveness of these heuristics, we select the same proportion (4%) of routers respectively according to each heuristic, and then we compare the statistical performance of overlay paths that can be constructed if the selected routers are upgraded to open routers. When open routers are partially deployed, it is possible that a number of endpoint-pairs cannot find any overlay paths constituted of the deployed open routers. To avoid interference of this systematic effect, we exclude all invalid endpoint-pairs in case of using each heuristic, and evaluate the statistics of overlay paths based on the intersection of the rest endpoint-pairs in every case, which in particular includes 40105 endpoint-pairs.

Figure 6 illustrates the statistics of overlay paths when open routers are deployed according to each heuristic, as well as the situations of direct IP paths and overlay paths constituted of end-host-relays. Clearly, the heuristic of RTD performs best. In terms of both E2E-delay and overlap, deploying open routers according to RTD is able to help overlay paths achieve better performance than both overlay routing without using open routers and direct IP routing, and not far from that obtained by



(a) Impact of E2E-delay



(b) Impact on Overlap

Figure 6. Evaluation of practical heuristics for deploying open routers

deploying open routers according to router significance. Note that we have shown above that the significance-based deployment of open routers is very close to the universal deployment and can be approximately taken as the optimal deployment when only a specific number of open routers are affordable to deploy.

Moreover, the cost of obtaining RTD can be safely reduced by performing `traceroute` just between a sample of all endpoint-pairs. For example, the results show that the correlation coefficient between the ranking vector of the observable router's RTDs based on all E2E-paths and that based on 50% E2E-paths is 0.96, and 0.71 even if only 5% E2E-paths are utilized to calculate the router's RTD. It implies that the RTD heuristic is practical to guide the selection of suitable deploying places for open routers.

Unfortunately, the other heuristics based on AS- or POP-level topological features cannot agreeably deploy open routers in suitable places to effectively enhance overlay routing's performance. This is mainly because these heuristics restrict open routers to a small set of ASes or POPs, which accordingly reduces the diversity of available overlay paths. A straightforward improving idea to these heuristics is setting an upper bound to the number of open routers that can be deployed in the same AS or POP, which will be studied in our future work.

VI. Conclusion and Future Work

This paper aims to arouse attention to the opportunity of leveraging novel Internet architectures to improve the E2E performance of multimedia communications. To this end, we perform a measurement study investigating whether and to what extent the universal or partial deployment of open routers would be able to benefit overlay routing systems that are designed to improve the transmission of multimedia data. Given that there have not been realistic open routers deployed in the current Internet, we develop a methodology to carry out the study with the currently available measurement approaches. Based on extensive topology and routing measurements collected between over 200 PlanetLab sites, our results show that even a small number of open routers, if deployed appropriately, will be able to significantly enhance overlay routing's ability of constructing low latency overlay paths for multimedia communications. Finally, we comparatively evaluate various heuristics for open router's placement, and suggest the router degree as an effective and practical heuristic for service providers to select suitable deploying places of open routers.

Interesting future work includes building more practical models that take an open router's capacity, weighted customer requirements, and commercial issues on ISP's cooperation and competition into account, as well as investigating how to improve open router's placement by integrating multiple heuristics.

Acknowledgment

We thank the reviewers for their detailed and helpful comments.

References

[1] "Cisco visual networking index - forecast and methodology, 2007-2012," June 2008.

[2] S. Tao, K. Xu, A. Estepa, T. Fei, L. Gao, R. Guerin, J. Kurose, D. Towsley, and Z.-L. Zhang, "Improving voip quality through path switching," in *Proc. IEEE INFOCOM'05*, March 2005.

[3] H. Zhang, L. Tang, and J. Li, "Impact of overlay routing on end-to-end delay," in *Proc. ICCCN'06*, October 2006.

[4] Y. Amir, C. Danilov, S. Goose, D. Hedqvist, and A. Terzis, "An overlay architecture for high-quality voip streams," *IEEE Trans. Mult.*, vol. 8, no. 6, pp. 1250–1262, 2006.

[5] S. Ren, L. Guo, and X. Zhang, "Asap: an as-aware peer-relay protocol for high quality voip," in *Proc. ICDCS'06*, 2006.

[6] S.-J. Lee, S. Banerjee, P. Sharma, P. Yalagandula, and S. Basu, "Bandwidth-aware routing in overlay networks," in *Proc IEEE INFOCOM'08*, 2008.

[7] J. Touch, Y.-S. Wang, L. Eggert, and G. Finn, "A virtual internet architecture," in *Proc. ACM SIGCOMM FDNA'03*, 2003.

[8] R. Gold, P. Gunningberg, and C. Tschudin, "A virtualized link layer with support for indirection," in *Proc. ACM SIGCOMM FDNA'04*, 2004.

[9] L. Peterson, S. Shenker, and J. Turner, "Overcoming the internet impasse through virtualization," in *Proc. HotNets'04*, 2004.

[10] J. Turner and D. Taylor, "Diversifying the internet," in *Proc. GLOBECOM'05*, 2005.

[11] "Open network laboratory," http://onl.arl.wustl.edu/.

[12] S. Neil, M. Ratul, and A. Thomas, "Quantifying the causes of path inflation," in *Proc. ACM SIGCOMM'03*, Karlsruhe, Germany, 2003.

[13] S. Savage, A. Collins, E. Hoffman, J. Snell, and T. Anderson, "The end-to-end effects of internet path selection," *ACM SIGCOMM Computer Communication Review*, vol. 29, no. 4, pp. 289–299, 1999.

[14] D. Andersen, H. Balakrishnan, M. Kaashoek, and R. Morris, "Resilient overlay networks," in *Proc. SOSP'01*, 2001.

[15] K. Gummadi, H. Madhyastha, S. Gribble, H. Levy, and D. Wetherall, "Improving the reliability of internet paths with one-hop source routing," in *Proc. OSDI'04*, 2004.

[16] T. Fei, S. Tao, L. Gao, R. Guerin, and Z. li Zhang, "Light-weight overlay path selection in a peer-to-peer environment," in *IEEE Global Internet'06*, 2006.

[17] M. Gunes and K. Sarac, "Resolving anonymous routers in internet topology measurement studies," in *Proc. IEEE INFOCOM'08*, 2008.

[18] H. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, and A. Venkataramani, "iplane: An information plane for distributed services," in *OSDI'06*, 2006.

[19] N. Spring, D. Wetherall, and T. Anderson, "Scriptroute: A facility for distributed internet measurement," in *Proc. USITS'03*, 2003.

[20] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, k. claffy, and G. Riley, "As relationships: inference and validation," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 1, pp. 29–40, 2007.

Li Tang is a Ph.D. candidate at Tsinghua University, China. He received his B.E. degree from Department of Automation at Tsinghua University in 2003. His research interests are in overlay networks, Internet routing and measurement, and large-scale distributed systems.

Yin Huai is senior undergraduate student in Department of Automation at Tsinghua University, China.

Jin Zhou is a researcher in Research Institute of Information Technology at Tsinghua University, China. He received his B.E. and Ph.D. degrees from Tsinghua University in 2000 and 2005 respectively. His research interests include peer-to-peer networks, network security, and multimedia streaming systems.

Hao Yin is an associate professor in Department of Computer Science at Tsinghua University, China. He received his B.E., M.E., and Ph.D. degrees from Huazhong University of Science and Technology in 1996, 1998, and 2002 respectively. His research interests include performance evaluation for the Internet and wireless network, image/video processing, multimedia over wireless network, and security.

Zhen Chen is an assistant professor in Research Institute of Information Technology at Tsinghua University, China. He received his B.E. and Ph.D. degrees from Xidian University China in 1998 and 2004 respectively. His research interests include network security and peer-to-peer systems.

Jun Li is a professor in Research Institute of Information Technology at Tsinghua University, China. He received his B.E. and M.E. degrees from Department of Automation at Tsinghua University in 1985 and 1988 respectively, and Ph.D. degree from Computer Science Department at New Jersey Institute of Technology in 1997. His research interests include network security, pattern recognition, and signal processing.

APPENDIX A. GEOGRAPHIC DISTRIBUTION OF PLANETLAB SITES COVERED BY OUR DATA SET

Figure 7. Geographic distribution of PlanetLab sites covered by our data set

APPENDIX B. PSEUDO-CODES FOR IDENTIFYING ALIAS CANDIDATES

Algorithm 1 routing symmetry heuristic

Input: *Paths*: All E2E paths between relevant endpoints
Output: *Aliases*: A list of candidate alias pairs

```
Aliases = set()
for (src,dst) in Paths.keys():
  if Paths.has_key((dst,src)):
    pathpair = [Paths[(src,dst)],Paths[(dst,src)
        ]]
    [shorterPath,reversePath] = sortByLength(
        pathpair)
    for i in range(1,len(shorterPath)-1):
      if (i <= len(shorterPath) / 2):
        Aliases.add((shorterPath[i],reversePath
            [-(i+1)]))
      else:
        Aliases.add((shorterPath[i],reversePath[
            len(shorterPath)-1-i]))
return Aliases
```

Algorithm 2 common predecessor and destination heuristic

Input: *Paths*: All E2E paths between relevant endpoints
Output: *Aliases*: A list of candidate alias sets

```
Aliases = set()
for (src,dst) in Paths.keys():
  for i in range(1,len(Paths[(src,dst)])):
    if (not Aliases.has_key((Paths[src,dst][i
        -1],dst))):
      Aliases[(Paths[src,dst][i-1],dst)] = set()
          ;
    Aliases[(Paths[src,dst][i-1],dst)].add(Paths
        [src,dst][i])
return Aliases.values()
```

APPENDIX C. DETERMINATION OF PROPER COMMON PREFIX LENGTH

To gain a deeper insight on the effect of N's value, we submit to *ally* all the candidate alias pairs identified by our heuristics to verify their validity. *Ally* uses three active-probing approaches (the source-address-based and IP-ID-based techniques as also used by iPlane plus another DNS-based one) to perform pairwise alias judgment, and is one of the most accurate methods for alias resolution at present. Among totally 32948 submitted candidate alias pairs, *ally* returns consistently explicit results to 14245 ones, but 'unknown' or conflicted[1] results to the others.

Varying the value of N and comparing the resulted alias pairs to the actual[2] ones affirmed by *ally*, Figure 8 illustrates how the false positive rate and false negative rate change respectively as the value of N increases from its minimum to maximum. To make a suitable balance, we let the parameter N, i.e., the length of common prefix, be 18 in the verification of those candidate alias pairs for which *ally* does not return consistent or explicit results.

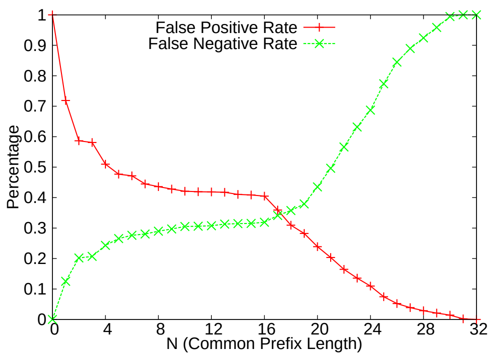


Figure 8. False positive and false negative rates under different values of N (common prefix length) in alias verification

APPENDIX D. NP-HARDNESS OF OODP PROBLEM

Theorem 1: The following optimal open router deployment problem (OODP) problem is NP-hard: given a network's topology represented by an undirected graph $G = (V, E)$, where the weight of edge $e = (u, v)$ is denoted by $w(u, v)$ and equivalent to the E2E-delay between $u \in V$ and $v \in V$, a set $P \subseteq V$ that is the set of all candidate routers, and a set $Q \subseteq V \times V$ of endpoint-pairs requiring overlay paths, find a set $R \subseteq P$ of k routers to deploy open routers such that it minimizes the given evaluation metric *Obj*$=\sum_{(x,y)\in Q} d(x, y)$, where $d(x, y) = \min_{r\in R}\{w(x, r)+w(r, y)\}$ is the smallest E2E-delay of the overlay paths constituted by the k deployed open routers between the endpoint-pair (x, y) . □

Proof: To prove the NP-hardness of OODP problem is equivalent to prove that the OODP decision problem (OODDP): if there exists a set $R \subseteq P$ such that $Obj \le B$, where B is a constant, is NP-complete.

First, OODDP is in **NP**, because the following verifier for OODDP runs in time polynomial in the size of Q : Verifier V($\langle G, P, Q, k\rangle$, $\langle R\rangle$)

If all the following conditions are true then accept else reject:

- $R \subseteq P$
- $|R| = k$
- *Obj*$=\sum_{(x,y)\in Q} d(x, y) \le B$

Next, we show that a well-known NP-complete problem *k-median problem* (KMP) [3] is reducible to OODDP. KMP can be summarized as follows: given a weighted undirected graph $\hat{G} = (\hat{V}, \hat{E})$, where the distance between $u \in V$ and $v \in V$ is denoted by $\hat{w}(u, v) \ge 0$, and two sets of vertices $M \subseteq \hat{V}$ and $N \subseteq \hat{V}$, determine whether there is a set of k nodes $L \subseteq M$ such that the objective function $\sum_{n\in N} \hat{d}(n) \le B$, where $\hat{d}(n) = \min_{l\in L}\{\hat{w}(n, l)\} \ge 0$ stands for the distance from n to its nearest node $l \in L$.

Define a function f that takes the input of a KMP instacne $\left\langle \hat{G}, M, N, k \right\rangle$, where $\hat{G} = (\hat{V}, \hat{E})$, and constructs an OODDP instance $\langle G, P, Q, k\rangle$, where $G = (V, E)$, as follows: for each node $m \in M$, create a node $p = f_1(m)$ in G, forming a set $P = \{f_1(m)|m \in M\}$; for each node $n \in N$, create two nodes $x = f_2(n)$ and $y = f_3(n)$, forming sets $X = \{f_2(n)|n \in N\}$, $Y = \{f_3(n)|n \in N\}$, and $Q = \{(f_2(n), f_3(n))|n \in N\}$; let $V = P \cup X \cup Y$, $E = \{(p, u)|p \in P, u \in X \cup Y\}$ and $w(p, u) = \frac{1}{2}\hat{w}(f_1^{-1}(p), f_i^{-1}(u))$ ($i = 2$ if $u \in X$, or 3 if $u \in Y$). f_1, f_2, and f_3 are all bijections. It is easy to find that the complexity of f is ploynomial. Now, we show that given a solution to KMP, we can accordingly find a solution to OODDP $\langle G, P, Q, k\rangle$, and vice versa.

Suppose L is a solution to KMP, and let $R = \{f_1(l)|l \in L\}$. Then $|R| = k$ and R is a solution to OODDP, because *Obj*$=\sum_{(x,y)\in Q} d(x, y) = \sum_{(x,y)\in Q} \min_{r\in R}\{w(x, r) + w(r, y)\} = \sum_{n\in N} \min_{l\in L}\{\hat{w}(n, l)\} = \sum_{n\in N} \hat{d}(n) \le B$.

Conversely, suppose R is a solution to OODDP, and $L = \{f_1^{-1}(r)|r \in R\}$. Then $|L| = k$ and it can be similarly proven that L is a solution to KMP.

Therefore, the same as the KMP problem, the OODDP problem should also be NP-complete, and accordingly the corresponding optimization problem OODP should be NP-hard. ■

[1] In a small portion (0.44%) of cases, *ally* returns contrary results when requested by two queries consisting of the same pair of aliases but in different orders.

[2] In fact, the alias pairs affirmed by *ally* are still not absolutely correct, but the result should be accurate enough to serve the purpose of selecting the best value of N.

[3] M. Garey and D. Johnson, *Computer and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman and Company, 1979.