Scalable and density-aware measurement strategies for overlay networks

Go Hasegawa* and Masayuki Murata*
*Graduate School of Information Science and Technology, Osaka University
1-3, Yamadaoka, Suita, Osaka 560-0871, JAPAN
Email: hasegawa@cmc.osaka-u.ac.jp

***Abstract*—In overlay networks, when we consider the effective and accurate measurement of underlay IP network between overlay nodes, it is important to take care the density of the overlay nodes in the network. In this paper, we propose the measurement strategy on the overlay networks which dramatically reduces the number of required measurement tasks for obtaining the up-to-date characteristics of full-mesh overlay paths. Our method does not require full-mesh information exchange between overlay nodes. One of the advantages of the propose method is that when the number of overlay nodes (N) increases our method does not require $O(N^2)$ measurement overhead, and the measurement overhead decreases when the density of the overlay nodes is larger than around 0.5. Through numerical evaluations, we show that our method can reduce the number of required measurements tasks by up to 1/50. We also find that we need to estimate the density of the overlay nodes in the network to determine the length of the measurement cycle for partial overlapping overlay paths.**

***Index Terms*—Overlay networks, network measurement, node density, measurement overlap**

I. Introduction

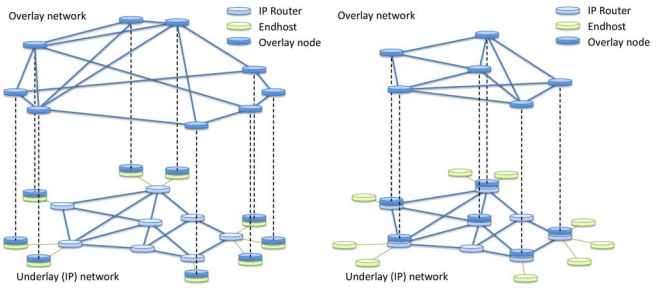
Overlay network, which is defined in this paper as an upper-layer logical network constructed upon the under-layer IP network, is now considered as an effective means to apply networked application services quickly. Typical examples are P2P-based applications including Skype [1], Grid, IP-VPN, and Application-Layer Multicast (ALM). Some of the overlay networks select an overlay-level route for data transmission according to network conditions such as link speed, delay, packet loss ratio, hop count, and TCP throughput between overlay nodes. For instance, in WinMX, an endhost can report the kind of network link used to connect to the Internet when joining the network. CDNs such as NetLightning [2] and Akamai [3] distribute overlay nodes (content servers) over the entire Internet and select appropriate source and destination hosts according to the network condition when the contents would be moved, duplicated or cached.

Furthermore, some overlay networks do not assume specific upper-layer applications and concentrate only on the routing of overlay network traffic. In Resilient Overlay Networks (RON) [4], each overlay node measures the end-to-end delay and packet loss ratio of the network path between the node and other nodes, and determines the route for the overlay network traffic originating from the node, which can be a direct route from the node to the destination node or a relayed route which passes through other node(s) before reaching the destination node. Thus, overlay routing can provide more effective traffic transmission compared to lower-layer IP routing. Furthermore, it can detect network failures (link and node failures, and misconfigured routing settings) and provide an alternate route in faster time than the IP routing convergence.

Due to its fundamental nature of overlay networks, it is a reasonable assumption that the characteristics of the overlay path between overlay nodes, such as IP-level route, latency, bandwidth-related information, packet loss ratio, and so on, is not known explicitly in advance. Therefore, for improving the performance of overlay networks, measuring overlay paths is an important task to obtain real-time and precise condition of overlay paths constructing the overlay network. Although some measurement mechanisms for overlay networks have been proposed in the previous works [4], most of them employs the full-mesh measurement, meaning that all of overlay paths between all possible node pairs would be monitored.

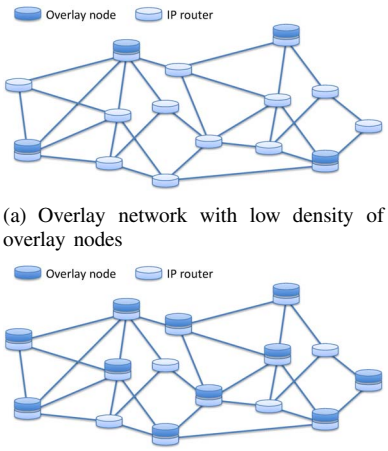
Those methods are effective for small-scale overlay networks by reducing the time required for obtaining enough information of overlay paths. For large-scale overlay networks, however, the increase of the measurement overhead and the decrease of the measurement accuracy due to path overlapping become a serious problem. For example, in RON, the number of participant overlay nodes is limited to around 50 [5]. To accommodate large-scale overlay networks, we need effective and scalable method for decreasing the measurement overhead. Furthermore, when the number of overlay nodes increases in the network, the measurement conflict [6] due to path overlapping arise as a big problem.

In this paper, we propose the measurement strategy on the overlay networks which dramatically reduces the number of required measurement tasks while obtains up-to-date characteristics of full-mesh overlay paths. Our mechanism is based on simple and existing technologies such as `traceroute`, and no full-mesh information exchange is required. We focus on the *density* of the overlay nodes, which is defined as the ratio of the number of overlay nodes to the number of IP routers, and construct scalable and density-aware measurement method. One attractive characteristics of the propose method is that when the number of overlay nodes (N) increases, our method does not require $O(N^2)$ measurement overhead, and the measurement overhead decreases when the density of the overlay nodes is larger than around 0.5. Furthermore, our mechanism does not require the knowledge of the density of the overlay nodes. To the best of our knowledge, this paper is the first one which focuses on the overlay node density for overlay network measurement.



(a) Endhost overlay network (b) Router overlay network

Fig. 1. Overlay network models



(a) Overlay network with low density of overlay nodes

(b) Overlay network with high density of overlay nodes

Fig. 2. Density of the overlay nodes

The remainder of this paper is organized as follows. In Section II, we introduce the overlay network model in this paper and explain some problems in measuring the overlay paths. We then propose the measurement strategies to alleviate one of the problems in Section III. In Section IV, we investigate the characteristics of overlapping of the overlay paths and show some direction for constructing the effective measurement strategies. Finally, Section V summarizes the conclusions of the present study and discusses areas of future consideration.

II. Network model and problem definition

A. Network model

We define *overlay network* as an upper-layer logical network constructed upon the under-layer IP network. Traditionally, overlay network services such as Skype is structured as an *endhost overlay* depicted in Figure 1(a), where the application program runs on endhosts connected to the IP network. In this case, the effectiveness of the overlay-level traffic routing, denoted as (*overlay routing*) in this paper, decreases since the overlay traffic cannot be routed inside the IP network. One possible way to resolve this problem is to introduce overlay nodes into the IP network, that is, a part of routers in the IP network behave as overlay nodes and participate the overlay network. We call this type of overlay network as *router-overlay*. In what follows, we focus on the router-overlay network environment. In addition, we omit the overlay nodes on the endhost, as depicted in Figure 1(b), since such overlay nodes do not contribute the overlay routing.

B. Density of overlay nodes and measurement requirement

We introduce the notion of *density* of the overlay nodes on the router-overlay network. The density is defined as the ratio of the number of overlay nodes to the number of routers in the network. For example, the density of the overlay network depicted in Figure 1(b) becomes 6/8=0.75.

Here we focus on the characteristics and performance of the network measurement (latency, packet loss ratio, and available bandwidth) of the overlay paths between overlay nodes on the overlay network with different density. We show the typical examples of overlay networks with low-density and high-density of the overlay nodes in Figures 2(a) and 2(b), respectively. When the overlay network has low density, the following characteristics can be found.

- Since the average hop count of overlay paths is large, the accuracy of network measurement is relatively small.
- The probability with which other overlay nodes exist on the route between two overlay nodes is small.
- The probability with which some overlay paths share the part of the route (path overlapping) is small.

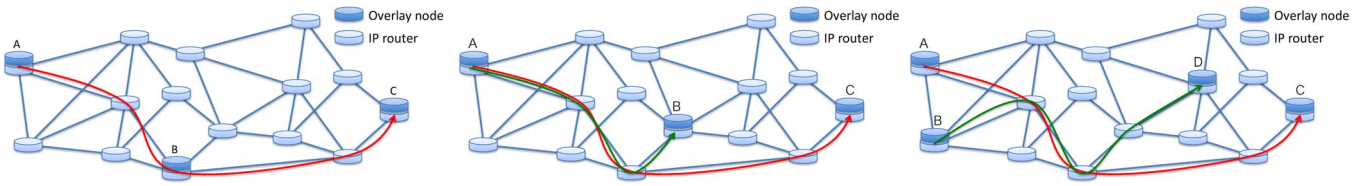
In this case, we can run lerger number of measurement tasks simultaneously without measurement conflict. Therefore, we can avoid the degradation of the accuracy of the measurement results. Furthermore, we can give longer time to each measurement task, which improve the measurement accuracy. We mean the term “measurement accuracy” as the accuracy of packet-interval-based bandwidth measurement methods, which degrades its accuracy when multiple measurement tasks are on the same path, or when the length of the measurement path is long.

On the other hand, when the density is high, we would have the following characteristics.

- Since the average hop count of overlay paths is small, the accuracy of network measurement is large.
- The probability with which other overlay nodes exist on the route between two overlay nodes is large.
- The probability with which some overlay paths share the part of the route is large.

In this case, when multiple measurement tasks are executed simultaneously, measurement conflict occurs frequently since many overlay paths share the IP-level route. Since the measurement conflict increases the network load by probe packets and degrades the measurement accuracy, we should avoid the measurement conflict.

There exists several researches to address the problem of measurement conflict [6, 7]. For example, in [7], the authors formulate the measurement conflict problem as a scheduling problem and propose a heuristic algorithm based on graph partitioning concepts. However, most of the previous works including [6, 7] assume that the overlay network has the knowledge of the AS-level and/or IP-level topologies and routing configuration of under-layer IP network. This assump-



(a) Complete overlapping (b) Half overlapping (c) Partial overlapping

Fig. 3. Classification of path overlapping and measurement conflict

tion is reasonable when the overlay network is applied to small-scale and manageable IP networks such as single ISP network. However, when the overlay network lays on larger-scale networks such as multi-ISP networks, the assumption cannot be satisfied.

Without the knowledge of the network topology and routing information, the overlay nodes should exchange the topology and routing information with other nodes for obtaining enough information on the under-layer network to avoid measurement conflict. However, the information exchange essentially requires $O(N^2)$ overhead, where N is the number of overlay nodes, since the degree of path overlapping and measurement conflict can be known only after the exchange of the topology and routing information. In what follows, we propose the measurement strategy which does not require $O(N^2)$ exchange of the topology and routing information.

C. Classification of path overlapping and measurement conflict

In this subsection, we classify the path overlapping in the overlay networks into three types and present key ideas to tackle the measurement conflict problems. We depict typical examples of path overlapping for each type in Figure 3.

1) Complete overlapping: Figure 3(a) shows the complete overlapping case. In this case, node B exists on the route between node A and node C, meaning that path AB and path BC are completely included in path AC. Therefore, when the measurement tasks on such paths are simultaneously executed, a measurement conflict occurs. Generally, we need full-mesh exchange of routing information to recognize this type of path overlapping.

In this paper, we apply the following simple idea to solve this problem: we omit the measurement of the longer path (path AC) and utilize the measurement results of shorter paths (path AB and BC) constructing the longer path to estimate the characteristics of the longer path. In the next section, we describe the detailed algorithm to realize this mechanism without full-mesh information exchange.

2) Half overlapping: Figure 3(b) depicts the example of half overlapping. In this case, path AB and path AC share the route from node A to the intermediate router. When the measurement tasks on those paths are executed at the same time, a measurement conflict occurs and the measurement accuracy degrades.

Note that node A can recognize this type of path overlapping by executing `traceroute` to nodes B and C, which can be done without information exchange with other overlay nodes. To avoid the measurement conflict in this case, node A should execute the measurement tasks on paths AB and AC *sequentially*.

3) Partial overlapping: Figure 3(b) shows the case of partial overlapping, where paths AC and BD share the route between two intermediate routers that are not source/destination overlay nodes. As in the case of complete overlapping, the full-mesh information exchange is required to become aware this type of path overlapping.

In this work, we avoid the measurement conflict in stochastic manner: for each overlay path, we estimate the maximum number of other overlay paths which are in partial overlapping relationships, and set the length of the measurement cycle according to the estimation results. That is, when the number of overlay paths in the partial overlapping relationship is larger, we set the longer measurement cycle for such paths to decrease the probability of occurring measurement conflicts. In Section IV, we investigate the characteristics of partial overlapping to obtain the guidelines to determine the length of measurement cycle.

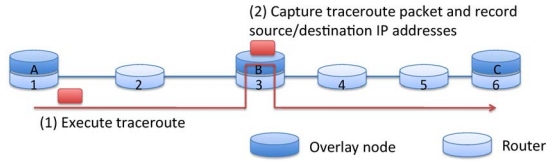
III. Reduction of Measurement Conflict

In this section, we propose the reduction method of the measurement conflict in the complete overlapping case explained in Subsection II-C 1) (Figure 3(a)). We assume that each overlay node knows the IP addresses of all other overlay nodes, and that the overlay node can execute `traceroute` command to other overlay nodes. Additionally, the overlay nodes can capture the `traceroute` packets passing through the overlay node, and record the source and destination IP addresses of corresponding `traceroute` command.

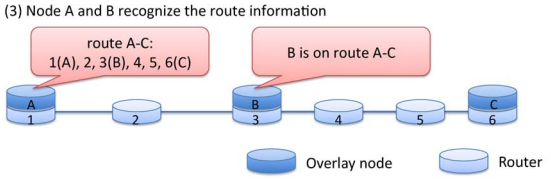
A. Proposed method

We summarize the algorithms to find the complete overlapping relationships and reduce corresponding measurement conflicts as follows.

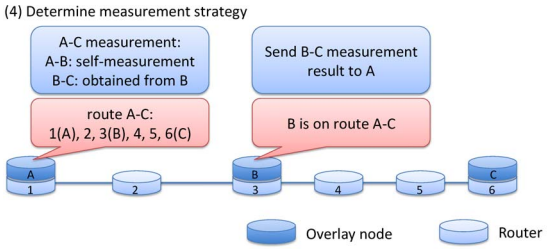
- Step 1: `traceroute` to an overlay node
  The source overlay node executes `traceroute` commands to the destination overlay nodes.
- Step 2: Capture the `traceroute` packets
  When a `traceroute` packet passes through an overlay node, the overlay node captures the packet and record the source and destination IP addresses of the `traceroute` command.
- Step 3: Check the complete overlapping
  Each overlay node summarizes the results of the execution of `traceroute` commands and captured other `the traceroute` packets, to check the existence of complete overlapping.



(a) Steps 1 and 2



(b) Step 3



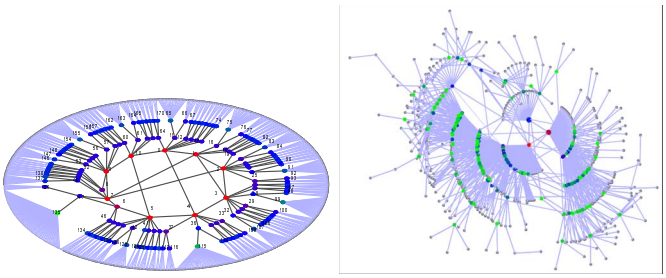
(c) Step 4

Fig. 4. Reduction method of measurement conflict

- Step 4-1: Determine the measurement strategy for source overlay node
  When there is no intermediate overlay nodes on the route to the destination overlay node, the overlay path between the source and destination overlay nodes is measured by the source node. When there is one or more intermediate overlay node, the overlay path is not measured directly, and the source node collects the measurement results of shorter paths constructing the overlay path, from the intermediate nodes. The source node also measures the path from the source node to the nearest intermediate overlay node.
- Step 4-2: Determine the measurement strategy for intermediate overlay node
  According to the results of the captured `traceroute` packets, the intermediate overlay nodes determine the source overlay node to which the measurement results are transferred.

Figure 4 depicts the example of the behavior of the proposed method. Node A executes a `traceroute` command to node C, and node B captures the packet and records the IP addresses of nodes A and C. Consequently, node A measures the path to node B, and the measurement results between nodes B and C is obtained from node B. Note that when there are two or more intermediate overlay nodes, we can apply the same algorithm in a recursive manner.

By this method, we can achieve the reduction of the required measurement tasks especially when the density of the overlay nodes in the network increases. That is, when the density increases, the probability with which the overlay path includes the intermediate overlay nodes on its route. Then the probability of measuring the path decreases.



(a) Abilene-inspired topology (b) BA topology

Fig. 5. Network topology for numerical evaluation

To estimate the measurement result of a longer path from the results of shorter paths constructing the longer path, we use the following simple calculation. Assuming that a longer path is constructed from shorter paths denoted as path 1, path 2, ... path N, and the measurement results of latency, available bandwidth, and packet loss ratio of each shorter path i are denoted by L_i, B_i, and p_i $(1 \leq i \leq N)$. Then, we estimate the measurement result of the longer path as follows:

- Latency: $L = L_1 + L_2 + ...L_N$
- Available bandwidth: $B = \min(B_1, B_2, ..., B_N)$
- Packet loss ratio: $p = 1 - (1 - p_1)(1 - p_2)...(1 - p_N)$

B. Numerical evaluation

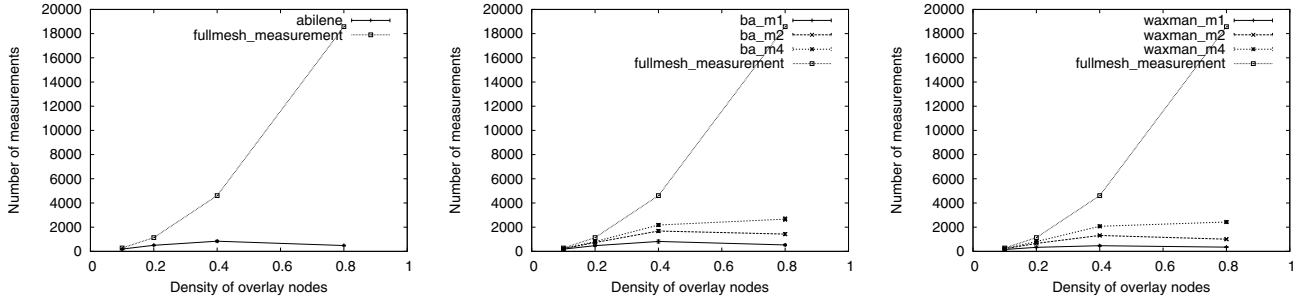
In this subsection, we evaluate the proposed method explained in the previous subsection. We used the following network topologies for under-layer IP networks.

Abilene-inspired topology
This network topology is proposed in [8], and has the similar structure to Abilene network. The topology has 171 routers and 178 links, and the distribution of the node degree (the number of outgoing links from each router) follows a power law. Figure 5(a) depicts the network topology.

BA topology
The generation algorithm of this network topology is proposed in [9]. The node degree distribution follows a power law. We generate the topology by using BRITE topology generator [10]. The number of routers is fixed to 171, and the number of links is m times as the number of routers, where m=1, 2, and 4. Figure 5(b) depicts the example of BA topology.

Random topology
This topology is based on the WAXMAN model [11] and is called as a random network. As in the BA topology, we generate the topology by using BRITE. The number of routers is fixed to 171, and the number of links is m times as the number of routers, where m=1, 2, and 4.

We evaluate the performance of the proposed method as follows. First, we place the overlay nodes on the routers in the network according to the density of the overlay nodes. Then we calculate the routes between all overlay node pairs by using Dijkstra's algorithm. We finally apply the proposed method



(a) Abilene-inspired topology (b) BA topology (c) Random topology

Fig. 6. Evaluation results of proposed method

and calculate the number of measurement tasks, including both directions, required to obtain the measurement results of all overlay node pairs. We compare the proposed method with a simple full-mesh measurement method where all paths between all overlay node pairs are measured directly. In the following results, we plot the average values and 95% confidence intervals of 1000 times iterations with different overlay node placements.

In what follows, we show the limited results due to space limitation. Note that we obtained similar results when we increased the number of routers and links in the IP network, and when we utilized the network topology of the actual ISP networks.

Figure 6 plots the number of measurement tasks required by the proposed method for obtaining the full-mesh measurement results in Abilene-inspired, BA, and Random topologies. We also plot the results of a simple full-mesh measurement method for comparison purpose. From Figure 6(a), we observe that when we use the Abilene-inspired topology, the number of required measurement tasks can be decreased up to 1/40 by using the proposed method. Furthermore, it is an interesting result that when the density of the overlay nodes is larger than around 0.5, the number of required measurement tasks decreases. This is because the probability with which the overlay path is directly measured rapidly decreases when the density increases. This results means that the proposed method is scalable to the density of the overlay nodes.

From Figures 6(b) and 6(c), we observe that the proposed method has similar characteristics on different network topologies. When m increases, that is, when the number of links in the network increases, the number of required measurement tasks increases with the proposed method. This is because when the number of links in the network increases, the routes between overlay nodes traverses various links, causing the decrease of the complete path overlapping. However, even with m=4, the proposed method can largely decrease the required measurement tasks especially when the density of the overlay nodes is large.

To investigate the effect of the network topology in more detail, we re-plot the evaluation results of the proposed method in Figure 7 when we use three network topologies with roughly equal number of nodes and links. We can see from this figure that Abilene-inspired topology and BA topology give the

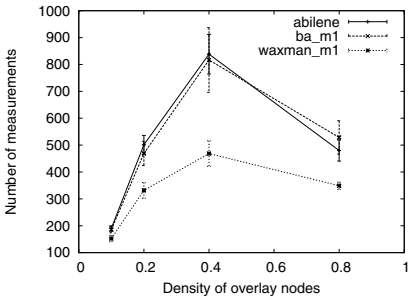


Fig. 7. Comparison results with three topologies

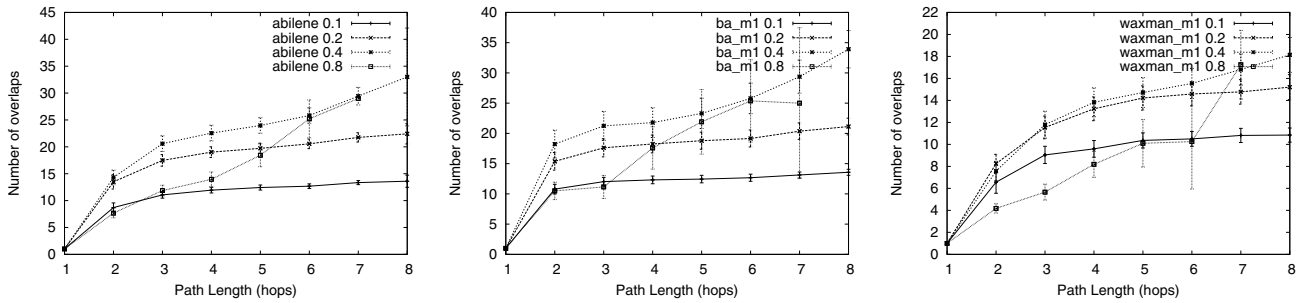
similar results, while random topology shows different results from the other two topologies. This is because of the difference in the distribution of node degree. In the power-law network topologies, there exists a few routers which has extremely large degree. Then, the overlay paths passing through such a high-degree router uses various links to arrive at and depart from the router. Therefore, the number of complete path overlapping decreases.

IV. Characteristics of partial overlapping

Finally, we investigate the occurrence of partial overlapping explained in Subsection II-C 3) (Figure 3(c)). In detail, for each overlay path, we calculate the number of other overlay paths in partial overlapping relationships. This metric affects the length of the measurement cycles for each overlay path to avoid measurement conflicts.

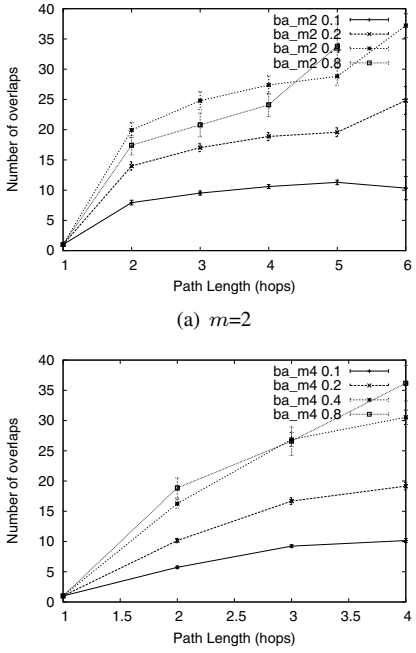
In Figure 8, we plot the number of overlay paths in partial overlapping relationships when we use three network topologies with m=1 and we change the density of the overlay nodes to 0.1, 0.2, 0.4 and 0.8. We calculate the average and 95% confidence intervals for overlay paths with equal hop count, and plot them as a function of hop count.

We observe from this figure that the number of partial overlapping paths is smaller than around 20-30 regardless of the topologies and overlay node density. We also note that the number of partial overlapping paths is not so correlated with the hop count of overlay paths, while the density of overlay nodes highly affects the number of partial overlapping paths. Furthermore, when the density of overlay nodes is 0.8, the



(a) Abilene-inspired topology (b) BA topology (m=1) (c) Random topology (m=1)

Fig. 8. Evaluation of partial overlapping (m=1)



(a) m=2

(b) m=4

Fig. 9. Evaluation of partial overlapping with BA topology (m=2, 4)

increasing trend as a function of hop count is quite different from those of small densities. Therefore, we conclude that to determine the length of the measurement cycle for partial overlapping overlay paths, we need to estimate the density of the overlay nodes in the network.

Figure 9 shows the results when we use m=2 and 4 for BA topology. We can see from this figure that the number of partial overlapping paths does not increase even when the number of links in the network increases. That is, we do not need to take care of the number of links in the network when we determine the length of the measurement cycle for each overlay path.

V. Conclusion

In this paper, we propose the measurement strategy for overlay networks, which is scalable to the density of the overlay nodes. We first classify the path overlapping patterns and propose the measurement strategies for each pattern. We then evaluate the reduction method of measurement conflicts for one of path overlapping patterns, and show that the proposed method can decrease the required measurement tasks up to 1/50. Furthermore, we evaluate the degree of the partial path overlapping and revealed that the number of partial overlapping paths is highly dependent on the density of the overlay nodes in the network.

For future work, we find appropriate metrics to estimate the number of partial overlapping paths and construct the measurement strategy for avoiding measurement conflict caused by partial overlapping. We then integrate the measurement strategies and evaluate the overall performance in terms of measurement accuracy, measurement overhead, and the degree of measurement conflicts.

Acknowledgment

This work was supported in part by the National Institute of Information and Communications Technology of Japan (NICT).

References

[1] Skype Web page, available at http://www.skype.com/.
[2] NetLightning Web Page, available at http://www.netli.com/services/netlightning/.
[3] Akamai Web Page, available at http://www.akamai.com/.
[4] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris, "Resilient overlay networks," in *Proceedings of 18th ACM Symposium on Operating Systems Principles*, Oct. 2001.
[5] A. Nakao, L. Peterson, and A. Bavier, "Scalable routing overlay networks," *ACM SIGOPS Operating Systems Review*, vol. 40, pp. 49–61, Jan. 2006.
[6] C. Tang and P. McKinley, "On the cost-quality tradeoff in topology-aware overlay path probing," in *Proceedings of ICNP 2003*, Nov. 2003.
[7] M. Fraiwan and G. Manimaran, "On the schedulability of measurement conflict in overlay networks," in *Proceedings of Networking 2007*, Nov. 2007.
[8] L. Li, D. Alderson, W. Willinger, and J. Doyle, "A first-principles approach to understanding the Internet's router-level topology," in *Proceedings of INFOCOM 2004*, Aug. 2004.
[9] A. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, Oct. 1999.
[10] BRITE: Boston university Representative Internet Topology gEnerator, available at http://www.cs.bu.edu/brite/.
[11] B. M. Waxman, "Routing of multipoint connections," *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 1617–1622, Dec. 1988.