



On Reducing Measurement Load on Control-Plane in Locating High Packet-Delay Variance Links for OpenFlow Networks

Nguyen Minh Tri¹(✉), Nguyen Viet Ha¹, Masahiro Shibata¹, Masato Tsuru¹,
and Akira Kawaguchi²

¹ Kyushu Institute of Technology, Kitakyushu, Japan
{tri.nguyen-minh414,nguyen.viet-ha503}@mail.kyutech.jp,
{shibata,tsuru}@cse.kyutech.ac.jp

² The City College of New York of The City University of New York, New York, USA
akawaguchi@ccny.cuny.edu

Abstract. We previously proposed a method to locate high packet-delay variance links for OpenFlow networks by probing multicast measurement packets along a designed route and by collecting flow-stats of the probe packets from selected OpenFlow switches (OFSs). It is worth noting that the packet-delay variance of a link is estimated based on arrival time intervals of probe packets without measuring delay times over the link. However, the previously used route scheme based on the shortest path tree may generate a probing route with many branches in a large network, resulting in many accesses to OFSs to locate all high delay variance links. In this paper, therefore, we apply an Eulerian cycle-based scheme which we previously developed, to control the number of branches in a multicast probing route. Our proposal can reduce the load on the control-plane (i.e., the number of accesses to OFSs) while maintaining an acceptable measurement accuracy with a light load on the data-plane. Additionally, the impacts of packet losses and correlated delays over links on those different types of loads are investigated. By comparing our proposal with the shortest path tree-based and the unicursal route schemes through numerical simulations, we evaluate the advantage of our proposal.

[AQ1](#)

1 Introduction

The OpenFlow technology was proposed more than a decade ago and is becoming widespread as a replacement solution for traditional network not only in data centers but also in enterprise networks and wide area networks. The ongoing prevalence of cloud computing and contents delivery networking requires flexible traffic engineering on a network connecting globally-distributed data-centers, which is often centrally managed by OpenFlow [1, 2]. By decoupling the

control-plane and data-plane, OpenFlow lets network operators configure, manage, monitor, secure, and optimize network resources very quickly via dynamic, automated programs. On the data-plane, switches forward packets based on per-flow rules installed and records the statistical information (flow-stats) of each flow. On the control-plane, a controller manages switches in the network by installing appropriate rules into each switch and collecting flow-stats from each switch.

Passive measurement by collecting per-link (from a physical input/output port of switch) traffic information via SNMP is commonly used to monitor and detect performance degraded links in traditional networks. However, in the edge-cloud computing for emerging IoT technologies, since a “link” between two nodes is not always physical but sometimes virtual, a per-link passive measurement cannot detect the performance degeneration of such virtual links. In OpenFlow network, by collecting the flow-stats from switches through the OpenFlow monitoring messages or by monitoring the OpenFlow-standard operating messages themselves, passive measurement approaches can operate in a per-flow manner without extra loads on the data-plane. However, there is a trade-off between the measurement accuracy and the load incurred on the control-plane, and thus some research efforts have been made. For example, [3] can calculate the network utilization by only using FlowRemoved and PacketIn messages of OpenFlow standard, but it cannot trace quickly changed links. In [4], the authors proposed a dynamic algorithm to balance the request frequency and accuracy.

Active measurement by probing packets is essential for flexibly and promptly monitoring any desired part of the entire network. The status of all links on a specific measurement route could be actively but aggregately monitored. However, probing at a high sending rate for a long duration can impose a greater load on the data-plane, and thus some research efforts have been made on how to reduce the load while still retaining reliability and precision. In [5], an infrastructure that focuses on reducing the flow entries and the number of probe packets in the round-trip time (RRT) monitoring is proposed. In [6], a measurement scheme that can cover all links in both directions while minimizing flow entries on switches is presented. For datacenter networks, an effective probe matrix is designed to locate real-time failures in [7]. However, those methods rely on unicast probing in an end-to-end (among servers or beacons) manner and generally suffer from a concentration of many overlapped probing paths traversing a small number of bottleneck links near the sender of the probing packets.

Based on those existing works, we previously presented a monitoring framework that combines an active measurement by probing multicast packets from a measurement host and a passive measurement by collecting the flow-stats from selective switches. Then, on that, we proposed a method to estimate the packet delay variance from the arrival time interval of packets at each switch and locate high packet-delay variance links. The variance of the packet delays on a link or on an end-to-end path can be clearly defined as a statistical value and can represent a degree of the packet delay variations or fluctuations. However, the term “packet delay variation” is sometimes related with jitter [8] and sometimes defined by

slightly different ways. We focus on the packet delay variance (the variance of the packet delays). Since links with a high packet delay variance are likely congested or physically unstable, it is of importance to monitor and locate such links in network performance management. Note that instead of directly measuring link delays and calculating delay variance, our method estimates the packet delay variance on a directional link or a directional segment between two ports (e.g., upper and lower ports of a link) based on the variances of packet arrival intervals monitored at each of those two ports. Differently from our previous work [9], the contribution of this paper is that we adopt an essential extension on a better route scheme [10] and provide an in-depth evaluation on how to reduce the load on the control-plane (i.e., the number of accesses to OFSs) while maintaining an acceptable measurement accuracy with a light load on the data-plane.

2 Monitoring Framework

2.1 Overview

The monitoring framework is based on that we previously proposed to monitor and locate high-loss links using multicast probing on OpenFlow networks [11]. It is for OpenFlow-based full-duplex networks consisting of the OpenFlow controller (OFC) and OpenFlow switches (OFS), with the measurement host (MH) that sends a series of multicast probe packets traversing all links in the network. An MH is directly connected to an OFS (called “measurement node”). The input port of the measurement node connecting the MH is called “root port”. Probing packets are launched at the MH toward the root port, traversed input ports of some OFSs, and finally discarded at some input ports (called “leaf port”). A measurement path from the root port to a leaf port is called “terminal path”. Note that the present method for estimation on delay variance requires an extension of flow entry and Flow-Stats Reply message to monitor the statistics of packet arrival time intervals on a specific flow [9].

First, as illustrated in Fig. 1, a measurement request from an MH is sent to the OFC. Then, the OFC obtains network topology, calculates probe packet routes, and installs them to OFSs. Probe packets are routed along a multicast tree route so that each probe packet passes through each link once and only once in each direction of each link separately to monitor bidirectional full-duplex links in both directions. The number of directed links on a terminal path is the terminal path length. A sequence of adjacent directed links along a path is called “segment” as a part of a terminal path. After that, the MH starts actively sending the probe packets. The packet arrival time interval at an individual input port on each OFS are passively recorded as flow-stats at each OFS and then, if needed, is collected by the OFC. Finally, the OFC calculates the packet-delay variance on a link (or a segment) between two ports and compared with a threshold to detect a high packet-delay variance link (or segment).

To reduce the loads on both the data-plane and the control-plane incurred by the measurement, two technical components, a flexible design of multicast measurement routes to cover all links in the active probing and a dynamic decision on

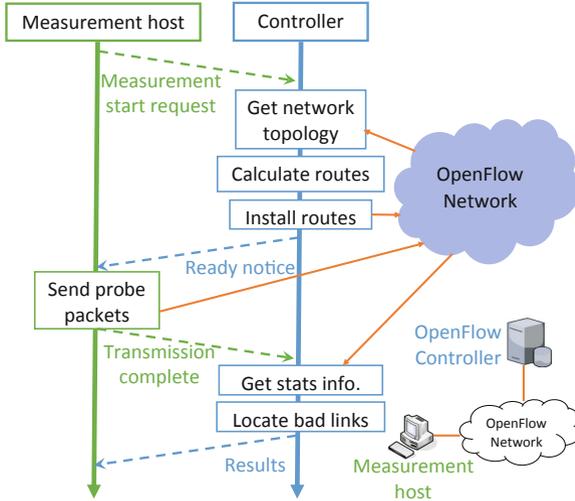


Fig. 1. Measurement process to locate performance degenerate links [11]

the sequential access order to switches for collecting the flow-stats, are required. In this paper, we focus on the former, i.e., the route scheme. The shorted-path tree based route scheme proposed in [9] suffers from generating many terminal paths so that it needs a large number of accesses to locate high packet-delay variance links. Therefore, as explained in the following subsection, we adopt a better route scheme based on [10] with fewer terminal paths to reduce the load on the control-plane while keeping an acceptable path length to minimize the load on the data-plane.

2.2 The Backbone-and-Branch Tree Route Scheme (BBT)

The proposed route scheme is called the backbone-and-branch tree route scheme (BBT). In this subsection, the BBT scheme is briefly explained as the example in Fig. 2. The Eulerian cycle algorithm is used to build backbone paths in the original undirected graph (network). Since an Eulerian cycle exists if and only if the graph consists of only even-degree vertices, first we need to remove all links between couples of odd-degree vertices (nodes) temporarily, called “omitted links”, see dashed lines in Fig. 2a. Then, we generate a backbone cycle by using the Eulerian cycle algorithm to cover all remaining undirected links. From the generated backbone cycle, we can build one or two backbone paths. To avoid too long terminal paths, the BBT T2 with two halves of the Eulerian cycle as backbone paths is used in this paper, see the bold lines in Fig. 2.

After building backbone paths, we divide each backbone path into multiple backbone segments with almost the same length. At the end node of each segment, called branch node, the reverse direction segment of route on the backbone path is added as extension of the route toward the measurement node, called

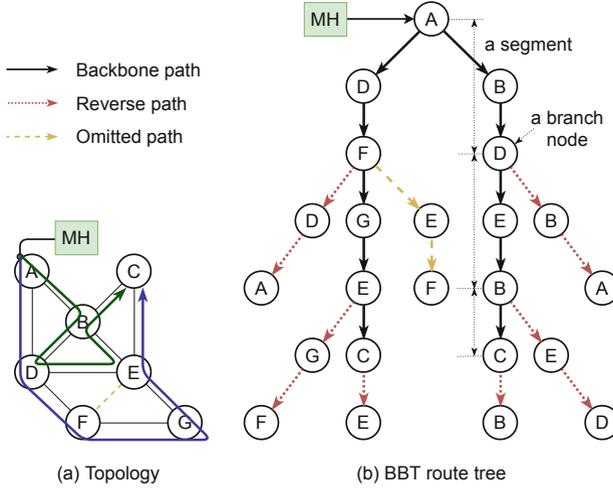


Fig. 2. Example of BBT route design

the reverse path, see the dotted lines in Fig. 2b. By this way, both directions of each full-duplex link should be traversed by a measurement path. The reverse path has the same length with its backbone segment but the opposite direction. Finally, we integrate additional paths of temporally omitted links into the route tree, see the dashed line in Fig. 2b. Those operations eventually construct a route tree consisting of multiple terminal paths for multicast measurements.

3 Estimate Packet Delay Variance from Arrival Intervals

To estimate the packet delay variance, a simple and direct method is measuring packet delay times of samples (i.e., probe packets in our case) and computing their unbiased variance. However, the packet delay time measurement requires matching and subtracting the arrival times of a same packet monitored at two different OFSs. Thus, the list of arrival times of all probing packets should be moved from a place to another, which induces a considerable load on the control and/or data planes. In our method, instead of direct measurement of per-packet delay times, each OFS monitors the arrival time intervals of two adjacent packets in a series of probe packets and computes their statistics to record locally and incrementally, which can be performed within each OFS independently and does neither require to store a long list of per-packet information nor to exchange it between OFSs or the OFC. After the above measurement of probe packets is finished, the OFC collects the arrival time interval statistics at each input port of OFSs and estimates the packet delay variance between two ports using the collected statistics by using an estimation method explained below.

The sequence diagram of probe packets is shown in Fig. 3. Let Q_k^j and D_{k+1}^j be the queuing delay at the k^{th} OFS of the j^{th} probe packet and the transmission

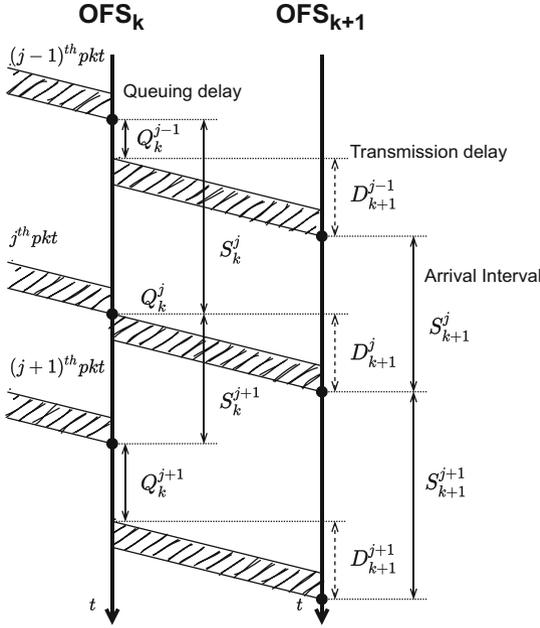


Fig. 3. Sequence diagram of packets probing

delay (including the propagation delay) of the j^{th} probe packet from the k^{th} OFS to the $(k+1)^{th}$ OFS, respectively. The arrival time interval S_{k+1}^j of the $(j-1)^{th}$ and j^{th} probe packets at the $(k+1)^{th}$ OFS is

$$S_{k+1}^j = S_k^j + (Q_k^j + D_{k+1}^j) - (Q_k^{j-1} + D_{k+1}^{j-1}) \quad (1)$$

where S_k^j is the arrival time interval of the $(j-1)^{th}$ and j^{th} probe packets at the k^{th} OFS. Note that, the arrival time interval S_1^j of j^{th} packet at the first OFS is equal to the initial sending time interval at the MH because we can assume no queuing delay between the MH and the first OFS.

Since we assume the bandwidth of each link and the probe packet size do not change in time, we have $D_{k+1}^j = D_{k+1}^{j-1}$, and thus the arrival time intervals of packets only depend on the queuing delays of OFSs (mainly at egress/output ports of OFSs) and the initial sending interval S_1^j .

$$S_{k+1}^j = S_k^j + Q_k^j - Q_k^{j-1} \quad (2)$$

The following ideal preconditions are defined to estimate the delay variance from the arrival time intervals. We will discuss the impact of a deviation from those conditions (i.e., correlated delays) later.

- Queuing delays of a packet at different links (different output ports) along the measurement path are independent within the measurement duration. Hence S_k^j and $Q_k^j - Q_k^{j-1}$ are independent.

- Queuing delays of succeeding packets at a link over time are independent and identically distributed within the measurement duration. Hence Q_k^j and Q_k^{j-1} are independent.

Therefore, the variance of the arrival intervals is expressed as follows

$$V[S_{k+1}] \cong V[S_k] + 2V[Q_k] \quad (3)$$

From Eq. 3, the queuing delay variance at the k^{th} OFS is

$$V[Q_k] \cong \frac{V[S_{k+1}] - V[S_k]}{2} \quad (4)$$

This means that, in general, the delay variance of a specific link or segment between two OFSs can be estimated from the difference of the arrival interval variances of those OFSs. Note that the arrival interval variance at each OFS is simply computed by

$$V[S_k] = E[(S_k)^2] - (E[S_k])^2 \quad (5)$$

where $E[(S_k)^2]$ and $(E[S_k])^2$ can be computed using the sample mean incrementally, that is, we do not need to store a long list of $\{S_k^j | j = 1, 2, \dots\}$.

One problem to consider is the packet loss. Possible holes in a series of probe packets due to packet losses should be considered and removed in the process of monitoring the arrival time intervals. If the j^{th} probe packet is lost somewhere and an OFS receives the $(j-1)^{th}$ and $(j+1)^{th}$ packets but not receive the j^{th} packet, then OFS discards the arrival time interval between $(j-1)^{th}$ and $(j+1)^{th}$ packets and does not count it in the statistics. To detect such holes by lost packets at OFS, the MH embeds a sequence number into ID field of IP header of each probe packet.

Another problem is called the ‘‘narrow interval’’; meaning that the j^{th} packet arrives at the k^{th} OFS before the $(j-1)^{th}$ packet departs from the OFS, i.e., two succeeding packets stay in the same queue. If two adjacent packets arrive at an OFS closely and meet similar congestion levels, Q_k^{j-1} and Q_k^j are similar and thus have a positive correlation. This problem will decrease our estimation’s accuracy on delay variance $V[Q_k]$. A simple solution is to enlarge the initial sending time interval at the MH, although it will prolong the measurement duration. In our simulation, we adopt this approach.

4 Locate High Packet Delay Variance Links

In the first step of the location process, the OFC queries OFSs that have leaf ports to collect the information on arrival time intervals at those ports and estimates the delay variance of each terminal path using the information at the leaf ports and the root port by (4). If the delay variances of all terminal paths are less than the threshold h , that means the network do not include any high delay variance link. Here, h is a design parameter that represents the target delay variance quality of links to maintain, which depends on the target applications.

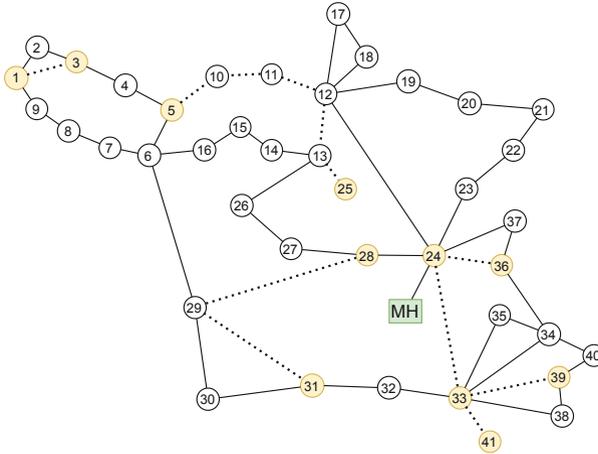


Fig. 5. Renater network topology

In the simulation, we compare the newly proposed route BBT T2 with the previous shortest path tree-based route (Model 2) and the unicursal route (the route has only one terminal path with a maximum length). Their information of terminal paths is in Table 1.

Table 1. Number of terminal paths and path lengths of route schemes

	Paths	Average	Min	Max
Unicursal	1	108	108	108
Model 2	26	5.5	3	12
BBT T2	8	19.6	8	28

Paths: Number of terminal paths
Average: Average length of terminal paths.
Min: The minimum length.
Max: The maximum length.

In the simulation, the parameters relating with packet delay times are set on each link as follows. A baseline static delay time of a link is set to a randomly selected fixed value from a range of $[10.0, 20.0]$ (ms). An additional dynamic delay (queuing delay) of each output port of OFS is a random variable with an exponential distribution that is independent of each other. The mean value (the expectation) of this random variable of dynamic delay is randomly selected from a range of

- $[5.0, 10.0]$ (ms) for each of a specific number of high delay variance links,
- $[2.0, 4.0]$ (ms) for each of 10% moderate delay variance links,
- $[0.2, 1.0]$ (ms) for each of other little delay variance links.

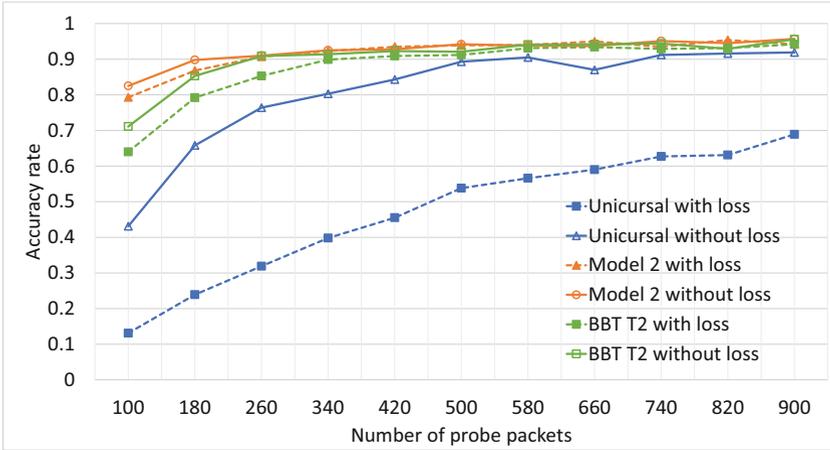


Fig. 6. Accuracy rate for locating 2 high-delay variance links with packet loss

We assume a random light loss rate in range of $[0, 0.01]$ on every link. The threshold h of high delay variance is 25. The initial sending interval of probe packets is 150 (ms) to avoid correlated delays among the adjacent packets and the narrow interval problem. The number of probe packets varies from 100 to 900. All resulting values are averaged over 1,000 measurement instances.

5.2 Simulation Results

Figure 6 shows the measurement accuracy depending on the number of probe packets. The measurement accuracy is defined as the ratio of the number of measurements in which all 2 high-delay variance links are correctly located to the total number of measurements (1,000 in our setting). We compare the results in two scenarios: with and without the packet loss. The estimation accuracy of delay variance relies on the number of probe packets. Therefore, the packet loss has a strong impact on the measurement accuracy. We see that a route scheme with longer terminal paths needs many probe packets to operate accurately. This is because each packet loss on an upstream link of measurement paths will make a hole of recorded arrival time interval on all remaining links, making the estimated value smaller or larger than the true value; these are “underestimation” or “overestimation”, respectively. The underestimation leads OFC to skip over high delay variance links. On the other hand, the overestimation leads OFC to unnecessarily and mistakenly seek high delay variance links. Additionally, accumulated errors over multiple links in a long segment will also create the underestimation or overestimation and result in a decrease of accuracy. The unicursal route suffers from a significantly low accuracy due to its long terminal path.

Figure 7 shows the number of the required accesses from OFC to OFSs until the high delay variance link location process ends in case of 2 high delay variance

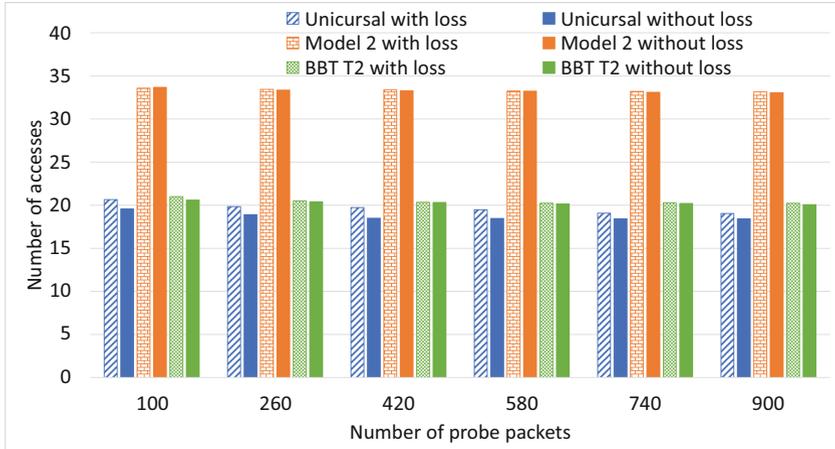


Fig. 7. Number of accesses for locating 2 high-delay variance links with packet loss

links, depending on the number of probe packets. Since the location process does not know the number of high delay variance links, the process lasts until it judges there is no other high delay variance links. Note that the results of the location process are not always correct because of errors in estimation. The shortest path tree-based route (Model 2) suffers from a larger number of terminal accesses due to a large number of terminal paths.

From results in Figs. 6 and 7, there is a trade-off between the load on the data-plane (the number of required probe packets for a certain estimation accuracy) and the load on the control-plane (required accesses). However, our proposed BBT T2 route can clearly balance the trade-off.

Although we assume the queuing delays on different ports along a terminal path are independent, a positive correlation may happen especially between high delay variance links. We examine this situation in case that 2 high delay variance links are positively correlated on the packet delays with the correlation coefficient $\rho = 0.5$ with the existence of packet losses. Figures 8 and 9 compare the performance with and without the correlation of delays between 2 high delay variance links.

From Fig. 8, the accuracy of the unicursal route (with a very long terminal path) is improved in the correlated case. This is because the positive delay correlation between ports within a segment makes the estimated value of the delay variance of the segment larger than the true value. This overestimation may introduce a fail-safe checking and increase the accuracy rate while increasing unnecessary accesses as shown in Fig. 9. Whereas, in Model 2 and BBT T2 with shorter terminal paths, the probability that these 2 corrected high delay variance links are positioned on the same terminal path is small. Therefore, the overestimation is small, and the results are similar both in the correlated case and the uncorrelated (independent) case as shown in Figs. 8 and 9. This suggests

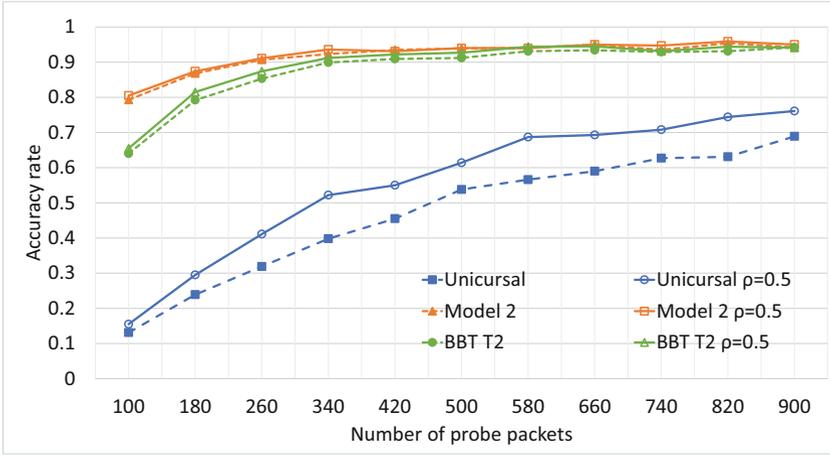


Fig. 8. Accuracy rate for locating 2 correlated high-delay variance links

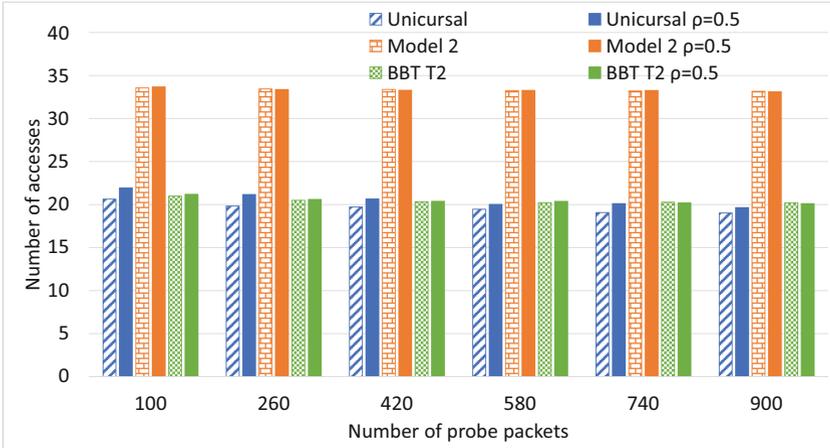


Fig. 9. Number of accesses for locating 2 correlated high-delay variance links

the resiliency of BBT T2 route scheme in correlated delays situations to some extent. As the result, in this simulation scenario with packet losses and correlated delays, BBT T2 scheme can locate 2 high delay variance links at an accuracy rate 0.9 with a small number (580) of the probe packets on the data-plane and a small number (20) of the switch accesses on the control-plane.

To have a detail view about the impact of packet losses and correlated delays, we investigate the locating process of the unicursal route with a long single terminal path. As above discussed, errors in estimation lead to the underestimation and overestimation. In the underestimation, the OFC may skip out high delay variance links because estimated values are smaller than true delay variances.

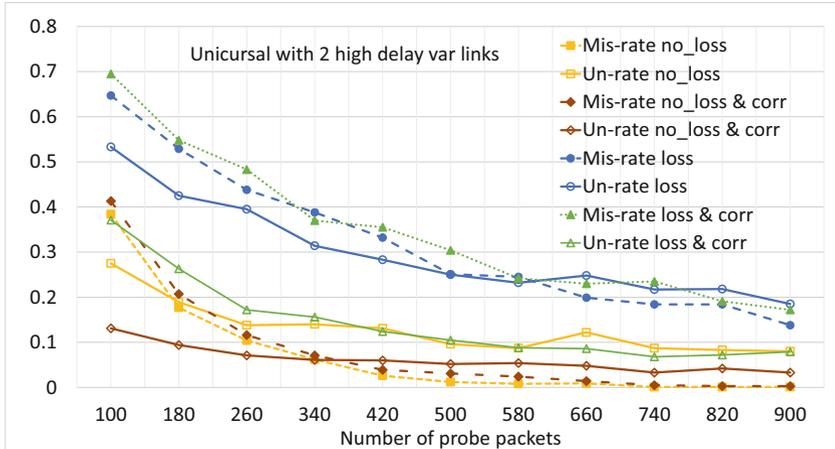


Fig. 10. Impact of the packet loss and delay correlation

On the other words, high delay variance links could be undetected. In the simulation, we count these situations and calculate the rate (called un-rate) over 1000 measurement instances. Similarly, “mis-rate” is the ratio of cases in which the OFC mistakenly detects a normal link into a high delay variance link to total measurements.

Figure 10 compares the error rates (un-rate and mis-rate) in four cases (with and without the packet loss and the positive delay correlation). With the packet loss, both mis-rate and un-rate are significantly increased, resulting in a low accuracy of measurement. About the impact of correlation, since correlated delays happen in 2 high delay variance links, the mis-rates are not changed with and without the correlation and enough small in case of no packet loss. On the other hand, because of overestimation on the delay variances, the un-rate is accidentally but significantly decreased in the correlation case regardless of the packet loss.

6 Concluding Remarks

Based on our previously proposed framework for monitoring and locating links with a high packet-delay variance in an OpenFlow network, in this paper, we apply an Eulerian cycle-based route scheme that can control the number of terminal paths and the lengths of them. Compared with the shortest path tree-based route in [9], the new proposal can reduce the number of accesses to switches by reducing the number of terminal paths while keeping a high accuracy rate by limiting the lengths of terminal paths. The proposed route scheme has been shown to be resilient with impacts of the packet losses and correlated delays through simulation results.

As a topic for future work, we will strive to adaptively optimize schemes for the multicast probe packet route by using the information on past measurement results to further reduce the number of accesses to OFSs.

Acknowledgements. These research results have been achieved by the “Resilient Edge Cloud Designed Network (19304),” NICT, and by JSPS KAKENHI JP20K11770, Japan.

References

1. Jain, S., Kumar, A., Mandal, S., et al.: B4: Experience with a globally-deployed software defined WAN. In: Proceedings ACM SIGCOMM 2013, pp. 3–14 (2013)
2. Hong, C.-Y., Kandula, S., Mahajan, R., et al.: Achieving high utilization with software-driven WAN. In: Proceedings ACM SIGCOMM 2013, pp. 15–26 (2013)
3. Yu, C., Lumezanu, C., Zhang, Y., et al.: FlowSense: monitoring network utilization with zero measurement cost. Lect. Notes Comput. Sci. **7799**, 31–41 (2013)
4. Chowdhury, S.R., Bari, M.F., Ahmed, R., Boutaba, R.: PayLess: a low cost network monitoring framework for software defined networks. In: Proceedings of 2014 IEEE NOMS, pp. 1–9 (2014)
5. Atary, A., Bremler-Barr, A.: Efficient round-trip time monitoring in OpenFlow networks. In: Proceedings of IEEE INFOCOM, pp. 1–9 (2016)
6. Shibuya, M., Tachibana, A., Hasegawa, T.: Efficient active measurement for monitoring link-by-link performance in OpenFlow networks. IEICE Trans. Commun. **E99B**(5), 1032–1040 (2016)
7. Peng, Y., Yang, J., Wu, C., et al.: deTector: a topology-aware monitoring system for data center networks. In: Proceedings of the 2017 USENIX Annual Technical Conference, pp. 55–68 (2017)
8. Demichelis, C., Chimento, P.: IP packet delay variation metric for IP performance metrics (IPPM). The Internet Engineering Task Force, IETF-RFC (2002)
9. Tri, N.M., Nagata, S., Tsuru, M.: Locating delay fluctuation-prone links by packet arrival intervals in openflow networks. In: Proceedings of the 20th Asia-Pacific Network Operations and Management Symposium, pp. 1–6 (2019)
10. Tri, N.M., Shibata, M., Tsuru, M.: Effective route scheme of multicast probing to locate high-loss links in OpenFlow networks. J. Inf. Process., 9 (2021)
11. Tri, N.M., Tsuru, M.: Locating deteriorated links by network-assisted multicast probing on OpenFlow networks. In: Proceedings of the 24th IEEE Symposium on Computers and Communications, pp. 1–6 (2019)
12. The Internet Topology Zoo, 14 May 2020. <http://www.topology-zoo.org/>